

STIC-Biotech/ChemLib

60964

From: Bunner, Bridget  
Sent: Monday, February 25, 2002 12:20 PM  
To: STIC-Biotech/ChemLib  
Subject: sequence search

Hi! I'd like a sequence search performed for case 09/755,017:

CRFE

1. the nucleic acid sequence of SEQ ID NO: 1
2. the nucleic acid sequence that encodes the amino acid sequence of SEQ ID NO: 2

thanks!

Bridget Bunner

Art Unit 1647  
CM1-10D12  
(703) 305-7148  
mailroom 10C01

Point of Contact:  
Toby Port  
Technical Info. Specialist  
CM1 ~~1E01~~ TEL: 308-3534

6104

	TYPE OF SEARCH:	VENDOR/COST(where applic.)
Searcher: _____	NA Sequences: <u>2</u>	STN: _____
Phone: _____	AA Sequences: _____	DIALOG: _____
Location: _____	Structures: _____	Questel/Orbit: _____
Date Picked Up: <u>2/25</u>	Bibliographic: _____	DRLink: _____
Date Completed: <u>2/26</u>	Litigation: _____	Lexis/Nexis: _____
Searcher Prep/Review: <u>10</u>	Full text: _____	Sequence Sys.: <u>es</u>
Clerical: _____	Patent Family: _____	WWW/Internet: _____
Online time: <u>10</u>	Other: _____	Other (specify): _____

CM 1101 1101 1101

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 26, 2002, 03:36:01 ; Search time 1214.49 Seconds

(without alignments)  
8334.794 Million cell updates/sec

Title: US-09-755-017-1

Perfect score: 942

Sequence: 1 atgaattggtaaatgacag.....tcttctaatacaagaataaa 942

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing First 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estoy:\*  
9: em\_hic:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_hic:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rtd:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	337.4	35.8	479	13	A0077154 CIT-HSP-2
C 2	203	21.5	523	10	A1604386 v74c09.x
C 3	201.8	21.4	692	12	A2090606 RPT-23-2
C 4	196.6	20.9	1394	12	AK017036 Mus muscu
C 5	196.6	20.9	3063	12	AK01560 Mus muscu
C 6	195.6	20.6	642	13	A2969227 2M0241J24
C 7	194.2	20.6	632	13	A2511623 1M0356C17
C 8	185.4	19.7	428	11	N68399 2A13B04.S1
C 9	184	19.5	326	13	A2694245 AST-2HBG3
C 10	180.4	19.2	788	13	BH11304 RPT-24-3
C 11	178.6	19.0	640	13	A2516219 RPT-11-3
C 12	175.6	18.6	588	13	A2413817 1M0188F05

C 13	172	18.3	740	13	A2607393
C 14	164	17.4	548	13	A2271125
C 15	162.6	17.3	632	13	A2765752 1M0562E19
C 16	159.6	16.9	680	13	A2709687
C 17	157.4	16.7	660	13	A2380178
C 18	156.2	16.6	675	13	A2396801 1M0161E05
C 19	154.8	16.4	1501	12	AK016338 Mus muscu
C 20	154.2	16.4	522	13	A2720443
C 21	153.8	16.3	646	13	A2638594
C 22	152.4	16.2	686	13	A2086625
C 23	152.2	16.2	534	13	A2312830
C 24	151.8	16.1	608	13	A2112932
C 25	151.8	16.1	797	13	A2909618
C 26	151	16.0	628	13	A2977433
C 27	150	15.9	508	13	A2373636
C 28	149.6	15.9	484	13	A0310400
C 29	149.4	15.9	632	13	A2392351
C 30	149.2	15.8	605	13	A2642411
C 31	148.4	15.8	680	13	A2235331
C 32	147.8	15.7	639	13	A2418543
C 33	147.8	15.7	796	11	BG197640
C 34	147.6	15.7	619	10	A1148854
C 35	147.4	15.6	532	13	A2519123
C 36	147.2	15.6	553	13	A0482125
C 37	147.2	15.6	614	13	A0453957
C 38	146.6	15.6	649	13	BH067870
C 39	146.4	15.5	534	13	A2664867
C 40	145.8	15.5	521	13	A2407823
C 41	145.8	15.5	678	13	BH067899
C 42	145.4	15.4	738	10	A1079550
C 43	145.4	15.4	762	11	BG193339
C 44	145.2	15.4	631	13	A2454523
C 45	145.2	15.4	797	13	A2885845

#### ALIGNMENTS

RESULT 1	A0077154	479 bp	DNA	GSS	20-AUG-1998
LOCUS	A0077154/C				
DEFINITION	CIT-HSP-2354D1.TF CIT-HSP Homo sapiens genomic clone 2354D1, DNA sequence.				
ACCESSION	A0077154				
VERSION	A0077154.1	GI:3438338			
KEYWORDS	GSS.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
AUTHORS	Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Sub,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.				
TITLE	Use of a random human BAC End Sequence Database for Sequence-Ready Map Building				
JOURNAL	Unpublished (1998)				
COMMENT	Other_GSSs: CIT-HSP-2354D1.TR				
	Contact: Mark Adams				
	Department of Eukaryotic Genomics				
	The Institute for Genomic Research				
	9712 Medical Center Dr., Rockville, MD 20850, USA				
	Tel: 301 838 0200				
	Fax: 301 838 0208				
	Email: mdadams@tigr.org				
	Closures are available from Research Genetics (info@resgen.com). BAC				
	end search page:				
	http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.				
	Seq primer: M13-21				
	Class: BAC ends.				
FEATURES	Location/Qualifiers				
source	1..479				
	/organism="Homo sapiens"				

Email: [mouseest@watson.wustl.edu](mailto:mouseest@watson.wustl.edu)  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.

RESULT	3		
LOCUS	AZ090606/c		
DEFINITION	AZ090606	692 bp	DNA
ACCESSION	RPCT-23-27B20	.TV	RPCT-23 Mus musculus genomic clone RPCT-23-27B20.
	AZ090606		
			GSS
			08-MAY-2000



VERSION	KEYWORDS
AZ090606.1	GI:7732649
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus. 1 (bases 1 to 692)
AUTHORS	Zhao,S., Niemman,W., Feldblum,T., Malek,J., Shatsman,S., Akintret ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Kroll,M., de Jong,P. and Fraser,C.M.
TITLE	Mouse BAC End Sequences from Library RPCI-23
JOURNAL	Unpublished (1999)
COMMENT	Other_GSSs: RPCI-23-27B20..TJ

Contact: Shanying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: [szhaoc@tigr.org](mailto:szhaoc@tigr.org)  
Clones are derived from the mouse BAC library RPc1-23. For BAC library availability, please contact Pieter de Jong ([pieterdejong.med.buffalo.edu](mailto:pieterdejong.med.buffalo.edu)). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea.ch Genetics ([inforesgen.com](http://inforesgen.com)). BAC end page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
Plate: 27 row: B column: 20  
Seq primer: T7  
Class: BAC ends.

FEATURES	Location/Qualifiers
source	1. .692

```

/organism="Mus musculus"
/strain="C57Bl/6J"
/db_xref="taxon:10090"
/clone="RPCT-23-27820"
/clone_lib="RPCT-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
ECORI; Site_2: EcoRI; Female C57Bl/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      205 a      164 c      163 g      160 t
ORIGIN

```

Query Match	21.4%;	Score 201.8;	DB 13;	Length 692;
Best Local Similarity	58.1%;	Pred. No. 1.2e-45;		
Matches 375;	Conservative 0;	Mismatches 267;	Indels 3;	Gaps 1

Qy	278	gtaacgcgagctcgtgtaagccagcttcacatcttcgcgccttgagggcaccgaatc	3377
Db	692	gctatgcggggatgcatagcacagttcttcataagacacttactacgagtggaactgagtgctg	6533
Qy	338	tctccctggccgtaatcctcttgataagtttgtagtatttgtcgccctcccatct	397
Db	632	tgcctccttggtgatggcctttccacacggcatgctgctgtgctgcctccactacataca	5733
Qy	398	cagttatcatcagccagagactcgcctcgcagttgtagagccgacatccttggttaactg	457
Db	572	ccacacttatatgcacaccccttcttcgcacatgacatggccatctcctcttgaggtggagagcc	5133
Qy	458	ttagtaactcagatgtgtgttctctacccctgactctccagctgcgcaactcttgacccctatg	517
Db	512	tttggaactctctgcacacgacagacatgcatcatgacacatgctctctgtaggcacatcacc	453
Qy	518	tgatagatcaactctctctgtggaagtcctccgcactgcgtcccaagtatctctgtgtgagcaa	5777
Db	452	tga-- -accactcttctctgtgataagtgctgtctccctggaagcctgcttgtagagacacag	396

QY	578	caggaaatgaggtcgaactctctcttgcaatggagctctccatcaataccctgaac	637
Db	395	TGGGAACAGAGACCTTACTTGTTTGGCGTGGAGCGTGAATTTTGGCTGCTGCTGAGAC	336
QY	638	lcaaccatatacatatgacttttaattgtcgcagacagatattaggaatcacagctctgaa	697
Db	335	TAAATTCTAGGACCTATGSCACAACTTGGCTCATGTCAGTGTGAACATCAAGTCAGATCTG	276
QY	698	gtcgacaaaagacatttggagacatgtggttcccatctaatgtggtgctcttttata	757
Db	275	GGCCGCAAAAGGCTCTGGGAGACTGTGGGTCCACACTTACTGTGGTTTCTTTTATG	216
QY	758	gtaaagccgtctctgtgttactgtcaacacacttgcgcagctccaaggaacgaagaa	817
Db	215	GCTGACGCAATGTACACGTATCTCGAGCTATCCACACGTAATCTGTGATGTAAGGAGAT	156
QY	818	tgtgtctctctctatgaatcatcttgaccacatgctgtaatcccatatataacta	877
Db	155	TTCGTGCCCCCTTTTATATCTATCTACTCCAAATGCTGAACCCCTGTGATTTATACCTTA	96
QY	878	ggaacaeaggagtaaggaagacattaaaggttgtgtcgaag	922
Db	95	GAACAAGAGATGTAAAGGGGCGCTGTGCAAGACTTGTAAAG	51
RESULT	4		
AKO17036			
LOCUS	AKO17036	1394 bp mRNA HTC	05-JUL-2001
DEFINITION		Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4933433E02, full insert sequence.	
ACCESSION	AKO17036		
VERSION	AKO17036.1	GI:12856091	
KEYWORDS		Cap trapper.	
SOURCE		Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone:14933433E02.	
ORGANISM		Mus musculus	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS		1 (bases 1 to 1394) Carninci,P. and Hayashizaki,Y.	
TITLE		High-efficiency full-length cDNA cloning	
JOURNAL		Methods in enzymology. 305, 19-44 (1999)	
PUBMED		99279253	
PUBMED		10349636	
AUTHORS		2 (bases 1 to 1394) Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.	
TITLE		Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	
JOURNAL		Genome research. 10 (10), 1617-1630 (2000)	
PUBMED		20499374	
PUBMED		11042159	
AUTHORS		3 (bases 1 to 1394) Shibata,K., Itoh,M., Aizawa,K., Nagoaka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,M., Sumi.N., Ishi.Y., Nakamura,S., Hazama,M., Nishize,T., Harada,A., Yamamoto,R., Matsunoto,H., Sakaguchi,S., Ikegami,T., Kasaiyagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Irawa,M., Ohara,E., Watabiki,M., Tameda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuzura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.	
TITLE		RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichipillary sequencer	
JOURNAL		Genome research. 10 (11), 1757-1771 (2000)	
PUBMED		20530913	
PUBMED		11076861	
AUTHORS		4 (bases 1 to 1394) The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.	
TITLE		Functional annotation of a full-length mouse cDNA collection	
JOURNAL		Nature 409, 685-690 (2001)	
PUBMED		5 (bases 1 to 1394)	
AUTHORS		Aachik,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,	





Db 1260 TCTGTTCTATACATTCCTACACAGCTGACACCCCTCATTTACAGCCTGAGAAAT 1319  
 QY 883 aagaggttaaggaagccttaaggttggtgcaaga 921  
 Db 1320 AAGGAGGTGATTGAGCTGTTAGAGAGTACTGGGAGA 1358

## RESULT 6

A2969227

LOCUS 642 bp DNA GSS 27-APR-2001  
 DEFINITION 2M0241J24R Mouse 10kb plasmid UUGC2M library Mus musculus genomic  
 clone UUGC2M0241J24 R, DNA sequence.

A2969227

ACCESSION A2969227  
 VERSION A2969227.1 GI:13840454  
 KEYWORDS GSS.

## SOURCE

ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

## REFERENCE

AUTHORS

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,

M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.

and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

## TITLE

COMMENT

- JOURNAL

COMMENT

Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: dunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0241 row: J column: 24

Seq primer: CACACGAGAACACGCTATGACC

Class: plasmid ends

High quality sequence stop: 642.

Location/Qualifiers

1. 642

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC2M0241J24"

/clone\_lib="Mouse 10kb plasmid UUGC2M library"

/sex="Female"

/lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (914732114191AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT

130 a 166 c 129 g 217 t

Query Match 20.8%; Score 195.6; DB 13; Length 642;  
 Best Local Similarity 56.5%; Pred. No. 6.2e-44;  
 Matches 363; Conservative 0; Mismatches 279; Indels 0; Gaps 0;

QY 233 tcccaaatgctagtaaatatgcaagcaatcaagaagtaacagtaacgtgctg 292  
 Db 1 tccccagatgatggtgcatcttctgtcagagagaaatcatttccgtatggagcctg 60  
 QY 293 tagccagcttctcaatattctgacctgagggtctctcaatattctccgcgcgtca 352  
 Db 61 tcaaccagctcttggatcatttcttgggtcctcagagagtggtccctcggcagcaa 120  
 QY 353 tgccttatagatgattgagctatttgcagcctctccatctacagtaacatcagcacc 412  
 Db 121 tggcatatgcatgattatgctatctgtaacccgttaagagtaactatattatgacaa 180  
 QY 413 agagactgctcagctcagctcagctcagctcagctcagctcagctcagctcagctc 472  
 Db 181 agcgtctgagcagctgagctgagcagctgagcagctgagcagctgagcagctgagc 240  
 QY 473 ggtgtctcctcagctcagctcagctcagctcagctcagctcagctcagctcagct 532  
 Db 241 tgcacacagcttggatcatttgcacacagcttggatcatttgcacacagcttggat 300  
 QY 533 tctgtaagctccctgagctcagctcagctcagctcagctcagctcagctcagctcag 592  
 Db 301 tctgtaagctccctgagctcagctcagctcagctcagctcagctcagctcagctcag 360  
 QY 593 aactatcctgctcagctcagctcagctcagctcagctcagctcagctcagctcagct 652  
 Db 361 ctttgctgctcagctcagctcagctcagctcagctcagctcagctcagctcagctc 420  
 QY 653 agctttatgctcagctcagctcagctcagctcagctcagctcagctcagctcagct 712  
 Db 421 acccttaccatgctcagctcagctcagctcagctcagctcagctcagctcagctcag 480  
 QY 713 tctggaacatgctcagctcagctcagctcagctcagctcagctcagctcagctcag 772  
 Db 481 tttccacagctcagctcagctcagctcagctcagctcagctcagctcagctcagct 540  
 QY 773 tctacatgcaacacacctcagctcagctcagctcagctcagctcagctcagctcag 832  
 Db 541 cgtatgctgagcagctcagctcagctcagctcagctcagctcagctcagctcagct 600  
 QY 833 atggaatcatgcaacacacctcagctcagctcagctcagctcagctcagctcagctcag 874  
 Db 601 atagcttctgacacacacctcagctcagctcagctcagctcagctcagctcagctcag 642

## RESULT 7

A2511623/c

LOCUS 632 bp DNA GSS 05-OCT-2000  
 DEFINITION 1M0356C17R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC1M0356C17 R, DNA sequence.

A2511623

ACCESSION A2511623  
 VERSION A2511623.1 GI:10692939  
 KEYWORDS GSS.

## SOURCE

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

## REFERENCE

AUTHORS

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,

M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.

and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

## TITLE

COMMENT

- JOURNAL

COMMENT

Contact: Robert B. Weiss

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: dunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0241 row: J column: 24

Seq primer: CACACGAGAACACGCTATGACC

Class: plasmid ends

High quality sequence stop: 642.

Location/Qualifiers

1. 642

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC2M0241J24"

/clone\_lib="Mouse 10kb plasmid UUGC2M library"

/sex="Female"

/lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (914732114191AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT

130 a 166 c 129 g 217 t

BASE COUNT

130 a 166 c 129 g 217 t



QY 875 tttagaacaaggagtgtaagaagctttaaaggttggttcgaagagctcttcaatca 934  
|||||  
DB 208 tttagaacaaggagtgtaagaagctttaaaggttggttcgaagagctcttcaatca 149  
QY 935 agaataa 942  
|||||  
DB 148 ACAATAA 141

## RESULT 9

A2694245

LOCUS 326 bp DNA GSS 18-DEC-2000  
DEFINITION AS1-2HG3005 Genetrap HL-60 Human Promyelocytic Leukemia Library  
Homo sapiens genomic 5' DNA sequence.

ACCESSION

A2694245

VERSION

A2694245.1

KEYWORDS

GSS.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 326)

Henkel, G., Llyanage, M., Pratt, E., Huang, D., Riley, M., Bernardino, A.,

Durick, K. and Pollok, B.

Exon-trap tags from a HL-60 Genomescreen(TM) Library

Unpublished (2000)

Contact: Greg Henkel

Gene Expression

Autera Biosciences Corp.

11010 Torreyana Road, San Diego, CA 92121, USA

Tel: 8584048436

Fax: 8584046719

Email: henkel@auterabio.com

Pools of cells were isolated from a Genomescreen(TM) library. The

library of cells was generated by retroviral integration of a gene

tagging element consisting of: 1) A promoterless beta-lactamase

proceeded by a splice acceptor as a reporter for gene expression;

2) A promoter driving neomycin resistance followed by a splice

donor to trap downstream exons. 3' RACE from neomycin gene was

performed using total RNA from isolated pools. Output was shotgun

cloned in pAMP-1 and used to transform DH5-alpha competent

bacteria. 5' ends of reported sequences were immediately preceded

by splice donor from the trapping construct.

Class: exon-trapped.

Location/Qualifiers

1..326

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="Genetrap HL-60 Human Promyelocytic Leukemia

Library"

/tissue\_type="acute promyelocytic leukemia"

/cell\_type="promyeloblast"

/cell\_line="HL-60"

/note="Organ: peripheral blood. Vector: pAMP-1; 3' RACE of

total RNA from genetrap pools: shotgun clone in pAMP-1 and

used to transform DH5-alpha competent bacteria."

BASE COUNT

109 a

65 c

57 g

95 t

ORIGIN

Query Match 19.5% Score 184; DB 13; Length 326;  
Best Local Similarity 100.0%; Pred. No. 9.1e-41;  
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 121 GAACAAGAGGATTAAGAGGCTTTAAAGGTTGCTGCAAGAGCTTCTATACAGAA 180  
QY 939 ataa 942  
|||||  
DB 181 ATAA 184

## RESULT 10

BH111304

LOCUS 788 bp DNA GSS 19-JUL-2001  
DEFINITION RPI-24-367N6.TJ RPI-24 Mus musculus genomic clone RPI-24-367N6,  
DNA sequence.

ACCESSION

BH111304

VERSION

BH111304.1

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 788)

Zhao, S., Mierman, W., Malek, J., Shatsman, S., Akimov, B., Levins, M.,

Tsegaye, G., Geer, K., Krol, M., Shwartsbeyn, A., Gebregorgis, E.,

Russell, D., de Jong, P. and Fraser, C.M.

Mouse BAC End Sequences from Library RPI-24

Unpublished (1999)

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPI-24. For BAC

library availability, please contact Pieter de Jong

(pdejong@tigr.org). Clones may be purchased from BACPAC

Resources (<http://www.chori.org/bacpac/orderingframe.html>). BAC endpage: [http://www.tigr.org/tigr/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tigr/bac_ends/mouse/bac_end_intro.html)

Plate: 367 row: N column: 6

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1..788

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPI-24-367N6"

/clone\_lib="RPI-24"

/sex="Male"

/cell\_type="Spleen/Brain"

/note="Vector: pTARBAC1. Site 1: BamHI; Site 2: BamHI. The

RPI-24 Mouse BAC library produced by Pieter de Jong. The

library was cloned in the pTARBAC1 cloning vector at the

BamHI sites using MhoI partially digested male C57BL/6J

DNA."

BASE COUNT

168 a

217 c

162 g

241 t

ORIGIN

Query Match 19.2% Score 180.4; DB 13; Length 788;  
Best Local Similarity 53.9%; Pred. No. 1.2e-39;  
Matches 413; Conservative 0; Mismatches 351; Indels 2; Gaps 2;

Db 126 TCCAAACATGTCCCAAGATGTTGTCCACCTTATTAACCAAGAAAGAACATCTCATTT 185  
 Oy 283 cgtgctgtgtgagccagcttcatatattctgctctggagctactgaatctctc 342  
 Db 186 GCCCATGATCAGACAGACATCTTGTATTTGGCCCTTTGGCGCTTGGAAGTCTGATTT 245  
 Oy 343 ctggcgtcatgtctctttagttagttgt- agcatctgtctgagctctccatctcagt 401  
 Db 246 TTGGACAGCATGCTCTATGAGAGAGTTGTGGGCATCTGCCACCCCTACACTACACTGT 305  
 Oy 402 taccatgacacagagactctgctccagcttgagcagcagcagctctggttactggttag 461  
 Db 306 CATCATGATGTTGAAAGATGTTGTGGCCCTGCTGACATCTTGATGATGATTTAG 365  
 Oy 462 taactagtgtgtgtctactccctgactctccagctctgacatctgtgacccctatgtat 521  
 Db 366 CTTCTGTGTGGACACACAAATTTCTCTGTAAGTTGCTTTTGTGGCCCAAGAAAT 425  
 Oy 522 agatcatctctctgtgaagtcctgacactgacatgacatctgtgtgtgagacacagc 581  
 Db 426 AAACACACTCTCTGTGTAATCCTGGCTGTCTCAAGCTGGCCCTGCCGACACTTTGAT 485  
 Oy 582 aaatgagctgaactattcctgtcagtgagctctccatcaatbaaccctgacactca- 640  
 Db 486 CAACCAATGCTCATCTAGCTGATGTTTGTCTTATGTGAGACCCCTTGTCTCAG 545  
 Oy 641 tccatatacatatgcttatttctgcagagatgagatgagatgagctctgaggtc 700  
 Db 546 TGTCTGATCTTACACGACATCTCTGAGCAATCTGAAATGCAAAAGAGGAGGCTC 605  
 Oy 701 gacaaaagcatttgagacatgtgttccatcaatgtgtgtctcttcttcttcttcttct 760  
 Db 606 GCAGAGAGGCTTTTCCACGCTGCTCCATCTCTGTGTGTGTGTGTGTGTGTGTGTGT 665  
 Oy 761 cagcgtctctgtgtactctgacacacactctgcccagctcccaagagaccaaagaatg 820  
 Db 666 TAGCCTATGTTGTTTATGTTATGTTGCTCCCTGACTGTGATCAGAGAGAGCAAGAAAT 725  
 Oy 821 ttctctctctatgagatcatctgacccatgctggaaccccttat 866  
 Db 726 TTTGTCACTGTTCCACAGCTTTTATCCCAAGATGATTCCTCAT 771

RESULT 11  
 AZ516219/c 640 bp DNA GSS 16-OCT-2000  
 LOCUS  
 DEFINITION RPCI-11-350623.TVK RPCI-11 Homo sapiens genomic clone  
 ACCESSION RPCI-11-350623, DNA sequence.  
 VERSION AZ516219  
 KEYWORDS  
 SOURCE GSS.  
 ORGANISM human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 640)  
 Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter  
 J.C.  
 BAC end sequences of library RPCI-11  
 Unpublished (1997)  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0200  
 Email: szhao@tigr.org  
 Clones are derived from the human BAC library RPCI-11. For BAC  
 library availability, please contact Pieter de Jong  
 (pieter@dejong.med.buflalo.edu). Clones may be purchased from  
 BACPAC Resources (http://bacpac.med.buflalo.edu/ordering) or from  
 Research Genet cs (info@resgen.com). BAC end search page:  
 http://www.tigr.org/tldb/humgen/bac\_end\_search/bac\_end\_search.html.

This BAC end was generated during the Rad process and may have  
 higher chance of clone tracking errors.  
 Seq primer: 77  
 Class: BAC ends.

FEATURES  
 source  
 1. 640  
 /organism="Homo sapiens"  
 /db\_xref="GDB:7634182"  
 /db\_xref="taxon:9606"  
 /clone="RPCI-11-350623"  
 /clone\_lib="RPCI-11"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /note="Vector: pBAC3.6; Site: 1: EcoRI; Site: 2: EcoRI;  
 RPCI11 Human Male BAC library"

BASE COUNT 185 a 99 c 140 g 216 t  
 ORIGIN

Query Match 19.0%; Score 178.6; DB 13; Length 640;  
 Best Local Similarity 93.0%; Pred. No. 3.5e-39;  
 Matches 198; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

Oy 731 atcaatgtgtgtctctcttcttcttcttcttcttcttcttcttcttcttcttcttctt 790  
 Db 640 ATCTATGTTGTGTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 581  
 Oy 791 cggccagctcccaagacaaagaatggttctctctcttcttcttcttcttcttcttctt 850  
 Db 580 CACCAGCTCCCAAGACGGGGGAAGATGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 521  
 Oy 851 tctgaatcccttatatactactagaaagaagagaaagagagagagagagagagagag 910  
 Db 520 TGTGAATCCCTTATATATACCTTAGAACAAGAGGTAAAGAGCTTTAAAGGT 461  
 Oy 911 tgggtgc-aagagctctcttaacaaagaat 942  
 Db 460 TGGTTCAGAAAGCTCTTCTTATCAAGAAATTA 428

RESULT 12  
 AZ413817 588 bp DNA GSS 03-OCT-2000  
 LOCUS  
 DEFINITION IM0188PF05 Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 ACCESSION AZ413817  
 VERSION AZ413817, GI:10537830  
 KEYWORDS  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 588)  
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
 M., Rose, M., Rose, R., Stokes, R., Tinger, A., von Niederhausern, A.  
 and Wright, D., Weiss, R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah  
 Km. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0188 row: F column: 05  
 Seq primer: GGTGTAAACGACGGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 588.

## FEATURES

SOURCE

Location/Qualifiers

1. 588

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUCG1M018F05"

/clone\_lib="Mouse 10kb plasmid UUCG1M library"

/sex="Male"

/lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource"

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT

157 a 139 c 104 g 188 t

Query Match

Best Local Similarity 18.6%; Score 175.6; DB 13; Length 588;

Matches 253; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

## BASE COUNT

157 a 139 c 104 g 188 t

Query Match

Best Local Similarity 18.6%; Score 175.6; DB 13; Length 588;

Matches 253; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

## FEATURES

SOURCE

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 740)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Rellly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A. and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0429 row: N column: 15

Seq primer: CACACAGAAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 740.

Location/Qualifiers

1. 740

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUCG1M0429N15"

/clone\_lib="Mouse 10kb plasmid UUCG1M library"

/sex="Male"

/lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource"

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT

276 a 134 c 159 g 171 t

Query Match

Best Local Similarity 18.3%; Score 172; DB 13; Length 740;

Matches 375; Conservative 0; Mismatches 320; Indels 1; Gaps 1;

## SOURCE

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 740)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Rellly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A. and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0429 row: N column: 15

Seq primer: CACACAGAAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 740.

Location/Qualifiers

1. 740

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUCG1M0429N15"

/clone\_lib="Mouse 10kb plasmid UUCG1M library"

/sex="Male"

/lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource"

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT

276 a 134 c 159 g 171 t

Query Match

Best Local Similarity 18.3%; Score 172; DB 13; Length 740;

Matches 375; Conservative 0; Mismatches 320; Indels 1; Gaps 1;

RESULT 13  
AZ607393/c 740 bp DNA GSS 13-DEC-2000  
LOCUS 1M0429N15R Mouse 10kb plasmid UUCG1M library Mus musculus genomic  
DEFINITION 1M0429N15R Mouse 10kb plasmid UUCG1M library Mus musculus genomic  
ACCESSION AZ607393  
VERSION AZ607393.1 GI:11729583  
KEYWORDS GSS.

house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 740)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Rellly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A. and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0429 row: N column: 15  
Seq primer: CACACAGAAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 740.  
Location/Qualifiers  
1. 740  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUCG1M0429N15"  
/clone\_lib="Mouse 10kb plasmid UUCG1M library"  
/sex="Male"  
/lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource"  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."



```

OY 304 ttcatattctgcttgaggagctactgaataatctccctgagccgtacgtcttgat 363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 517 TTCACCTTCTGATTTTATATAGATGCTGAGTGTGCTGCTGCGAGTGTGATTTGAT 458
OY 364 aggtttgaagctattgtcgcctctccatctactgaatcatatcatcaccagagactcgc 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 457 CGATATTAAGGCATTAAGTAAACCCCTTTGTATGAGATGACATGTCAGGAAGTGTGC 398
OY 424 ctccagcttgagcagccgcacccctcgggttactggttttagtaactcagtggtgtcctacc 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 357 TTCACATTAATTAAGTGGTGTATTTATCTGTGCGATTAAGAGATGCTTTGATACATACAA 338
OY 484 ctgactccacagctgcacactgtgacccctatgtgagatagatcactctctctgtgaagtc 543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 337 TTGAATTCACACTTATGTTTCTGTGCGTGAATGAGATTAATATCTTCTCTGTGATATT 278
OY 544 cctgacagctcgaatgatactctgtgttgagacaacagacaatgaggtcgaactatcctt 603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 277 CCTCCAGTGTAGTACTGCTGTGTCAGACACAGTCAATGTTTACTCATATTCACT 218
OY 604 gtcaatgagctctccatctaatatccctcgacatccatctatataatgctttat 663
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 217 GTTTTGTGTTTATTAAGTGAACGACACATCTCAAGATTTCTTCTACTGTTACATC 158
OY 664 gtccagagagatltgagatacagctcgtgagagtcgacaaagacattgagacatgt 723
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 157 ATCTATCATGATTTCTGAAGATCAGTCTGTGCTGCGAGGATTAAGCTTCTCTACTGT 98
OY 724 gttccatctaatgtgtgtctctttttagtaacagccgtctctgtgactgcaa 783
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 97 ACATACACACCTGACGCTGTTGCAATTTTCAAGAACATGCTCTCTCATGTA-TTCCCG 39
OY 784 ccaactcgccagctccaagcaagcaagaagatg 819
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 38 CCAAGTCTCTCTACTCCTTACCTGATCAAGATTAAGTG 3

```

```

RESULT 14
LOCUS A2271125 548 bp DNA GSS 26-JUL-2000
DEFINITION RPCI-23-146J4-TV RPCI-23 Mus musculus genomic clone RPCI-23-146J4,
DNA sequence.
ACCESSION A2271125
VERSION A2271125.1 GI:9484742
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 548)
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akimov,
B., Levins,M., McGinn,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
unpublished (1999)
Other GSSs: RPCI-23-146J4-TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pleter@dejong.med.bufile.edu). Clones may be purchased from
BACpac Resources (http://bacpac.med.bufile.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 146 row: J column: 4
Seq primer: T7
Class: BAC ends.
Location/Qualifiers

```

FEATURES

```

source
1..548
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-146J4"
/clone_11b="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site:1:
EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBAC3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies).
BASE COUNT 147 a 124 c 134 g 143 t
ORIGIN

```

```

Query Match 17.4% Score 164; DB 13; Length 548;
Best Local Similarity 61.9%; Pred. No. 4.1e-35;
Matches 260; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

```

```

OY 517 gtgataagatcattctctgtgaagtcctgacgtcgaatgatactctgttgagaca 576
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 470 GTAATAGACCATTTCTCTGTGAGTCCAGCCCTTTGAACATGACTGCACTGACACA 411
OY 577 acagcaaatgagctgaactatctctgtcagtgagctctccatctaataccctgaca 636
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 410 AGTGTGAATGAAGCTGACGCTTAATGTTTGGAGCTTTGCTCTCTGCTGCTCAAC 351
OY 637 ctcacatctatacatatgcttttctgtgcgagagatltgagatacagctctgta 696
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 350 CTGATCTGCGGTAACCTATGTTGCTCAAGCAAGTCAAACTCCGTTCTGCTGAG 291
OY 697 ggtcgacaaagaacatlttgagacatgtgttccatctaatgtgtgtctctttat 756
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 290 AGTCGCGGGAAGGACATTTAATACGTGCTTCAACATCTGCTGTGCTGCCCTTCAT 231
OY 757 agtaacgctctctgtgactgcaaccactctgcgcagctccaagacaaagaag 816
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 230 TTCACAGTATCAGATGATGATGCTGACCTCCCTCAAGCTCACTCAAGAAAGAAA 171
OY 817 atggttctctctctatgaaatcttgacccaatgctgaatcccttatatacact 876
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 170 ATCATGCTCTGTTCTATGACATGTCACACCTACCTCAACCATTCATCTACCTTG 111
OY 877 aggaacaagagagtaagaagagctttaaaggttggtgcaagagctctcttaacag 936
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 110 AGGAATTAAGCATTTAAGGCTCCCTCGAGAGGCGACTCAACAAGAGTTTGGGTCAG 51

```

```

RESULT 15
LOCUS A2765752 632 bp DNA GSS 16-FEB-2001
DEFINITION IM0562E19R Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0562E19 R, DNA sequence.
ACCESSION A2765752
VERSION A2765752.1 GI:12882100
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 632)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
Plasmid Inserts
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss

```



Tue Feb 26 09:22:08 2002

us-09-755-017-1.rst



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 26, 2002, 03:41:12 ; Search time 1356.19 seconds

(without alignments)  
11458.824 Million cell updates/sec

Title: US-09-755-017-1

Perfect score: 942

Sequence: 1 atgaattgggttaatgacag.....tcttcttaatacaagaataa 942

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_da:\*
- 2: gb\_hlg:\*
- 3: gb\_in:\*
- 4: gb\_ov:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_om:\*
- 20: em\_ov:\*
- 21: em\_ov:\*
- 22: em\_pat:\*
- 23: em\_ph:\*
- 24: em\_pl:\*
- 25: em\_ro:\*
- 26: em\_sts:\*
- 27: em\_sy:\*
- 28: em\_un:\*
- 29: em\_vi:\*
- 30: em\_htgo\_hum:\*
- 31: em\_htgo\_inv:\*
- 32: em\_htgo\_rnd:\*
- 33: em\_htg\_hum:\*
- 34: em\_htg\_inv:\*
- 35: em\_htg\_rnd:\*
- 36: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	940.4	99.8	44788	AL133267	AL133267 Human DNA
2	940.4	99.8	176277	AC025941	AC025941 Homo sapi
3	727.8	77.3	1074	HS302593	AJ302593 Homo sapi
4	726.2	77.1	1074	HS302584	AJ302584 Homo sapi
5	726.2	77.1	1074	HS302585	AJ302585 Homo sapi
6	726.2	77.1	1074	HS302586	AJ302586 Homo sapi
7	726.2	77.1	1074	HS302587	AJ302587 Homo sapi
8	726.2	77.1	1074	HS302588	AJ302588 Homo sapi
9	726.2	77.1	1074	HS302589	AJ302589 Homo sapi
10	726.2	77.1	1074	HS302590	AJ302590 Homo sapi
11	726.2	77.1	1074	HS302591	AJ302591 Homo sapi
12	726.2	77.1	1074	HS302592	AJ302592 Homo sapi
13	726.2	77.1	100375	HS193B12	Z98744 Human DNA
14	726.2	77.1	166758	AC024428	AC024428 Homo sapi
15	726.2	77.1	176277	AC025941	AC025941 Homo sapi
16	665.2	70.6	185523	AL589742	AL589742 Mus muscu
17	665.2	70.6	222824	AL589651	AL589651 Mus muscu
18	647.4	68.7	648	U86270	U86270 Homo sapien
19	630.6	66.9	648	U86271	U86271 Homo sapien
20	607	64.4	646	U86275	U86275 Homo sapien
21	599.6	63.7	222824	AL589651	AL589651 Mus muscu
22	588.4	62.5	1214	RATOLIRCE	L34074 Rat OLI rec
23	538.8	57.2	942	HS302546	AJ302546 Homo sapi
24	537.2	57.0	942	HS302537	AJ302537 Homo sapi
25	537.2	57.0	942	HS302538	AJ302538 Homo sapi
26	537.2	57.0	942	HS302539	AJ302539 Homo sapi
27	537.2	57.0	942	HS302540	AJ302540 Homo sapi
28	537.2	57.0	942	HS302541	AJ302541 Homo sapi
29	537.2	57.0	942	HS302542	AJ302542 Homo sapi
30	537.2	57.0	942	HS302543	AJ302543 Homo sapi
31	537.2	57.0	942	HS302544	AJ302544 Homo sapi
32	537.2	57.0	942	HS302545	AJ302545 Homo sapi
33	537.2	57.0	942	HS80119	AL022727 Human DNA
34	397.2	42.2	152311	AC011571	AC011571 Homo sapi
35	393.2	41.7	130279	HS031316	AL121944 Human DNA
36	380	40.3	185302	AL359352	AL359352 Mus muscu
37	380	40.3	251546	AL365336	AL365336 Mus muscu
38	370	39.3	178295	AL390860	AL390860 Human DNA
39	370	39.3	180657	AC091612	AC091612 Homo sapi
40	368.4	39.1	190889	AL357039	AL357039 Homo sapi
41	365.8	38.8	939	HS302565	AJ302565 Homo sapi
42	365.8	38.8	939	HS302566	AJ302566 Homo sapi
43	364.2	38.7	939	HS302555	AJ302555 Homo sapi
44	364.2	38.7	939	HS302560	AJ302560 Homo sapi
45	364.2	38.7	939	HS302561	AJ302561 Homo sapi

## ALIGNMENTS

RESULT 1  
LOCUS AL133267  
DEFINITION Human DNA sequence from clone RP3-408B20 on chromosome 6 Contains ESTs, STSs and GSSs. Contains a gene and two pseudogenes for novel 7 transmembrane receptors (Olfactory family) and a gene for a novel protein similar to 60S acidic ribosomal protein p2 (RPLp2), complete sequence.

ACCESSION AL133267  
VERSION AL133267.9 GI:10185396  
KEYWORDS HTG; 7 transmembrane; Olfactory receptor; ribosomal protein; RPLP2.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 44788)  
AUTHORS Williams, S.  
TITLE Direct Submision  
JOURNAL Submitted (25-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone

## COMMENT

requests: clonerequest@sanger.ac.uk  
 On Sep 18, 2000 this sequence version replaced gl:9588470.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence has been finished according to sequence map criteria  
 as follows. An attempt is made to resolve all sequencing problems,  
 such as compressions and repeats, but not necessarily within known  
 annotated human repeat sequence elements (e.g. Alu). Where the  
 sequence is ambiguous, there is an annotation using the 'unsure'  
 feature key.

The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information  
 on the WORMPEP database can be found at  
<http://www.sanger.ac.uk/projects/C-elegans/wormpep> This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping  
 Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr6>

REPEAT: This sequence is not the entire insert of clone  
 RP3-408B20 is from the library RPci-3 constructed at the Roswell  
 Park Cancer Institute by the group of Pieter de Jong. For further  
 details see <http://Dacpac.med.buffalo.edu/>  
 VECTOR: pcypac2

IMPORTANT: This sequence is not the entire insert of clone  
 RP3-408B20 it may be shorter because we sequence overlapping  
 sections only once, except for a 100 base overlap.  
 The true left end of clone RP1-31316 is at 44689 in this sequence.  
 The true right end of clone RP1-193B12 is at 100 in this sequence.

## FEATURES

## source

1. 44788

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="6"

/clone="RP3-408B20"

/clone.lib="RPci-3"

/complement(1..100)

/note="match: STS: Em:HS193B12"

222..535

/note="AluX repeat: matches 3..301 of consensus"

932..1870

/gene="dJ408B20.1"

932..1870

/gene="dJ408B20.1"

/note="dJ408B20.1 (novel 7 transmembrane receptor  
 (olfactory family) (hs6M1-33p) pseudogene)"

match: CDNAS: Em:M64392 Em:AF102516 Em:Y15524 Em:Y15525  
 Em:X89699 Em:AF102521 Em:AF042016 Em:AF044033 Em:AF102525  
 Em:AF044034 Em:AF044035 Em:AJ003145 Em:Y14442 Em:AF044037  
 Em:Y10529 Em:AF102529 Em:AF044038 Em:AF01565 Em:Y07557  
 Em:X89706 Em:AF042020 Em:M64391 Em:AF102532 Em:AF102533  
 Em:M64376 Em:AF102535 Em:M64377 Em:M64378 Em:AF102537  
 Em:AF091573 Em:AF091574 Em:AF091577 Em:AF091578  
 Em:AF102540 Em:M64385 Em:AF044053 Em:M64386 Em:M64387  
 Em:M64388 Em:AF091580 Em:DI2820

match: proteins: Tr:O76000 Tr:O76001 Tr:O76002 Tr:O35434  
 Sw:Q15062 Tr:O9WV09 Sw:Q13607 Tr:O90806 Tr:O63394  
 Tr:O9WV13 Tr:O9WV18 Tr:O9T084 Tr:O95006 Tr:O9Y389  
 Tr:O9Q222 Sw:O95156 Sw:O95157 Tr:O9UGF6 Tr:O9Y299  
 Tr:O9UJ15 Tr:O62007 Sw:P23275 Tr:O95371 Tr:O9UJ40  
 Tr:O9Q218 Tr:O9Q219 Tr:O9PS3 Tr:O9Y256 Tr:O9Q221"

/codon\_start=1

/pseudo

/evidence="not\_experimental"

/complement(1518..1963)

/note="match: GSS: Em:AQ456128"

2024..2298

/note="Alu repeat: matches 1..280 of consensus"

2299..2382

/note="42 copies 2 mer ag 79% conserved"

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

2559..2650

/note="2 copies 46 mer 91% conserved"

/complement(4227..4576)

/note="match: STS: Em:G31214"

4719..5074

/note="L2 repeat: matches 2375..2750 of consensus"

/complement(4834..5686)

/note="match: GSS: Em:AQ748194"

5098..5236

/note="L2 repeat: matches 1381..1525 of consensus"

5707..5843

/note="MIR repeat: matches 102..240 of consensus"

6275..6509

/note="L1M4 repeat: matches 5093..5329 of consensus"

/complement(6590..6889)

/note="match: GSS: Em:U94441"

6679..6752

/note="MER76 repeat: matches 613..685 of consensus"

6753..6850

/note="L1P repeat: matches 1..98 of consensus"

6820..7010

/note="L1P repeat: matches 580..776 of consensus"

7006..7615

/note="L1P repeat: matches 900..1509 of consensus"

7611..9145

/note="L1P repeat: matches 3813..5344 of consensus"

9751..10411

/note="HERV repeat: matches 3244..3919 of consensus"

10535..10563

/note="LTR16A repeat: matches 203..307 of consensus"

10964..11131

/note="MLT1H repeat: matches 1..168 of consensus"

11705..12141

/note="L1M4 repeat: matches 3898..4367 of consensus"

13143..13516

/note="match: GSS: Em:AQ100659"

13668..13803

/note="MIR repeat: matches 58..199 of consensus"

complement(15225..15506)

/note="match: GSS: Em:AQ100750"

15627..16228

/note="match: GSS: Em:AQ544878"

16824..17089

/note="AluYb repeat: matches 29..311 of consensus"

17090..17158

/note="AluYb repeat: matches 1..81 of consensus"

17533..17848

/note="AluX repeat: matches 3..298 of consensus"

17789..18240

/note="match: STS: Em:G31213"

20021..20249

/note="L1P repeat: matches 5904..6155 of consensus"

20708..21649

/gene="dJ408B20.2"

20708..21649

/note="match: CDNAS: Em:AF044034 Em:Y14442 Em:X89700  
 Em:Y07557 Em:AF042023 Em:AF091573 Em:AF091574 Em:X89670  
 Em:M64391 Em:AF102540 Em:AF102543 Em:M64388 Em:DI2820  
 Em:AF034896

match: ESTs: Em:M68399

match: proteins: Tr:O43883 Tr:O9Q221 Tr:O921K0 Tr:O9Q220  
 Tr:O63394 Tr:O95499 Tr:O9Q219 Sw:P23270 Tr:O9Y387  
 Tr:O9Q222 Tr:O9Y299 Tr:O62943 Tr:O90807 Tr:O90808  
 Tr:O9WV13 Tr:O63395 Tr:O9Y918 Sw:P23266 Sw:P37069  
 Sw:O95156 Tr:O9UGF4 Tr:O9W086 Tr:O60403 Tr:O76000  
 Sw:P37070 Tr:O9QW38 Tr:O77758 Tr:O9Q217 Tr:O9R0K3  
 Tr:O9R0K4 Tr:O60412 Tr:O9Q700 Tr:O9QW36 Tr:O9UJ40  
 Sw:P37072 Sw:P37071 Tr:O9Y385"

/codon\_start=1

/evidence="not\_experimental"

/product="dJ408B20.2 (novel 7 transmembrane receptor  
 (olfactory family) (hs6M1-32))"

```

/protein_id="CAC14158.1"
/db_xref="gi:10944516"
/translation="MMNWNDSIDIEFILLGRSDRPMLEFLLVFLVLSYTYTIFGNLT
IIIVSRIDTKLHTPMYFFLLISLIDCYTTCVPMQVNVNCSIRKVISYRGVADLE
IFLALGTEYLILLAVMSDFRPAICRPLHSVIMHORICLOLAASAVTGSNSVMS
TTLQPLCDPBYVDHFLCEAFVPLKLSCHVSTANAEFLVSLFHLIPLTLLISY
AFIVRAVLRISAGROKRCFSGSHLIVASFSTAVSYVLPSPSSKQKQWVSL
FYGIAPMINDPLIYTLINKKEGFKRLVARVFLIKK"
repeat_region
/notes="Vtr15 repeat: matches 127. 212 of consensus"
misc_feature
/notes="complement(21244..21671)"
repeat_region
/notes="match: GSS: Em:A0077154"
21964..21999
/notes="18 copies 2 mer ca 100% conserved"
repeat_region
23289..23944
/notes="LIMB2 repeat: matches 5508. 6168 of consensus"
24195..24291
/notes="FLAM_C repeat: matches 38. 136 of consensus"
24448..24781
/notes="match: GSS: Em:B75689"
24537..24857
/notes="match: GSS: Em:A0102724"
24739..25051
/notes="Alusq repeat: matches 1. 310 of consensus"
repeat_region
25407..25944
/notes="LIM4 repeat: matches 5206. 5763 of consensus"
25945..26050
/notes="Alusg/x repeat: matches 195. 302 of consensus"
26055..26632
/notes="match: GSS: Em:A0057601"
26069..26126
/notes="LIMB7 repeat: matches 6111. 6169 of consensus"
26164..26220
/notes="LIMB7 repeat: matches 6111. 6168 of consensus"
26221..26391
/notes="Alub repeat: matches 137. 306 of consensus"
repeat_region
Query Match 99.8%; Score 940.4; DB 9; Length 44788;
Best Local Similarity 99.9%; Pred. No. 1.4e-258;
Matches 941; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 481 accctgactctcagctgcgaactctgtgaccctatgtgatagatcacttctctgtaa 540
|||||
Db 21168 ACCCTGACTCTCCACTGCACTCTGTGACCCCTATGATGATGATCTTCTGTGAA 21247
OY 541 gtccctgactgctcaagtatctctgtgtgagacaagacaatgagctgaactatc 600
|||||
Db 21248 GTCCCTGCACTGCTCAAGTATCTGTGTGAGACAACAGCAATGAGCTCACTATTC 21307
OY 601 ctgtcagtgagctcttcacatacaccctgacactcaccatcattatcatatgcttt 660
|||||
Db 21308 CTGTGACAGGAGCTCTTCATCTAATACCCCGACACATCATCTATATCATATCTTTT 21367
OY 661 attgtccgagcagatitgagatagacagctgtgagagctgcgaagaacatttgagca 720
|||||
Db 21368 ATTGTCCGAGCAGATTTGAGGATACAGTCTGCTGAGAGTCCGACAAAGCATTTGGACA 21427
OY 721 tgtgttcccatcctaattgtgtgtctctttttataglacagccgtctctgtactgt 780
|||||
Db 21428 TGTGTTCCCATCTAATGTGTGTCTCTTTTATAGACAGCGCTCTGTGTACCTG 21487
OY 781 caaccacacttgcagctccagctccaggaaccaagaagaatggtctctctctcatgatac 840
|||||
Db 21488 CAACCACTTCGCCAGCTCCAGCAAGACCAAGAAAGATGTTCTCTCTATGTGAATC 21547
OY 841 attgcacccatgctgaatcccttatatacacttaagaaagaggttaagaagagc 900
|||||
Db 21548 ATTGCACCCATCTGAATCCCTTATATATACACTTTAGAACACAGAGGTAAGAGAGGC 21607
OY 901 tttaaaaggttggttgcaagagctctcttaacaaagaataa 942
|||||
Db 21608 TTAAAGGTGTGTGCAAGAGCTCTTATATCAAGAAATAA 21649
RESULT 2
AC025941 176277 bp DNA HTG 26-MAY-2000
LOCUS Homo sapiens chromosome 6 clone RP11-635011 map 6, WORKING DRAFT
DEFINITION AC025941.2 GI:7408057
VERSION AC025941.2 GI:7408057
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 176277)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 6, clone RP11-635011
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 176277)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouhagalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Chovel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArillano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lien,C., Liu,G., Locke,K., MacDonald,P., Margulis,N.,
McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McNeeters,R.,
Meldrum,J., Meneses,L., Mihova,T., Miranda,C., Mlenaga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Tagilias,J.,
Testaye,S., Theodore,J., Threlk,A., Travers,M., Trillano,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (17-MAR-2000) Whitehead Institute/MIT Center for Genome

```

## COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 4, 2000 this sequence version replaced g1:7259782.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

## Project Information

Center project name: L8577

Center clone name: 635\_O\_11

## Summary Statistics

Sequencing vector: M13: M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 168929 bases at least Q40

Consensus quality: 172827 bases at least Q20

Consensus quality: 174151 bases at least Q20

Insert size: 176000; agarose-fp

Quality coverage: 4.5 in Q20 bases; agarose-fp

Quality coverage: 4.6 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 15 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1      2997: contig of 2997 bp in length
*
*      2998 3097: gap of 100 bp
*
*      3098 3097: gap of 100 bp
*
*      5607 5706: gap of 100 bp
*
*      5707 5706: gap of 100 bp
*
*      9209: contig of 3503 bp in length
*
*      9210 9309: gap of 100 bp
*
*      9310 13614: contig of 4305 bp in length
*
*      13615 13714: gap of 100 bp
*
*      13715 20639: contig of 6925 bp in length
*
*      20640 20739: gap of 100 bp
*
*      20740 26545: contig of 5806 bp in length
*
*      26546 26645: gap of 100 bp
*
*      26646 37068: contig of 10423 bp in length
*
*      37069 37168: gap of 100 bp
*
*      37169 48083: contig of 10915 bp in length
*
*      48084 48183: gap of 100 bp
*
*      48184 59015: contig of 10832 bp in length
*
*      59016 59115: gap of 100 bp
*
*      59116 70420: contig of 11305 bp in length
*
*      70421 70520: gap of 100 bp
*
*      70521 80466: contig of 9946 bp in length
*
*      80467 80566: gap of 100 bp
*
*      80567 92325: contig of 11759 bp in length
*
*      92326 92425: gap of 100 bp
*
*      92426 110572: contig of 18147 bp in length
*
*      110573 110672: gap of 100 bp
*
*      110673 131898: contig of 21226 bp in length
*
*      131899 131998: gap of 100 bp
*
*      131999 176277: contig of 44279 bp in length.

```

## FEATURES

## Source

```

1..176277
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6"
/clone="RP11-635011"
/clone_lib="RPC1-11 Human Male BAC"
1..2997
/misc_feature
/feature="assembly_fragment"
3098..5606
/misc_feature

```

```

/misc_feature
/feature="assembly_fragment"
5707..5706
vector_side:left"
/misc_feature
/feature="assembly_fragment"
9310..13614
/feature="assembly_fragment"
13715..20639
/feature="assembly_fragment"
20740..26545
/feature="assembly_fragment"
clone_end:77
vector_side:right"
/misc_feature
/feature="assembly_fragment"
37169..48083
/feature="assembly_fragment"
48184..59015
/feature="assembly_fragment"
59116..70420
/feature="assembly_fragment"
70521..80466
/feature="assembly_fragment"
80567..92325
/feature="assembly_fragment"
92426..110572
/feature="assembly_fragment"
110673..131898
/feature="assembly_fragment"
131999..176277
/misc_feature
/feature="assembly_fragment"
BASE COUNT 54493 a 33357 c 33254 g 53772 t 1401 others
ORIGIN

```

```

Query Match 99.8%; Score 940.4; DB 2; Length 176277;
Best Local Similarity 99.9%; Pred. No. 1.6e-258;
Matches 941; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 atgaattggttaaagaacgacatcacagaggttatctctcgtggttcacagatca 60
|||||
DB 154174 ATGAATGGGTAAAGACGACATACAGAGGTTATCTGGGTTTCACAGATCA 154115
QY 61 ccttgctgagcttccacacctctgtgtctcttgatcttcaactgtgacatctt 120
|||||
DB 154114 CCTTGCTGAGATTTCCACTCTCTTGCTGCTCTTGATTTCTTACACTGTGACATCTTT 154055
QY 121 ggcacatcagacatattctagctgacgctgagacacccaactcacaatcccatg 180
|||||
DB 154054 GCGAATCTGACCATTTACTAGTGCACGCGACACCAACTTCATACCCCATGTTAT 153995
QY 181 tttttcttaccacatctacactctctgactctgttctgttaccacatgtaagtcacaaa 240
|||||
DB 153994 TTTTTCATATTTCTGGCCTTGGGCGCTACTGAAATCTTCTCGCGCGTCAATGCTTTT 153815
QY 241 atgcctagtaaatctatctgacgacatcagaaagaatcaagttatcgtgctgtgtagccag 300
|||||
DB 153934 ATGCGTAGTAATTTATGCGACATCAGGAAGTAATCAGTTATCGCGCTGTATACCCAG 153875
QY 301 ctttcacatctctgagccttggggctactgataatctctctcgtgacgtgacgtctt 360
|||||
DB 153874 CTTTTCATATTTCTGGCCTTGGGCGCTACTGAAATCTTCTCGCGCGTCAATGCTTTT 153815
QY 361 gataagttgtgactatctgtgctctccatctcctcaatgatatgacacagagacac 420
|||||
DB 153814 GATAGGTTGTGATGATTTGTCGCTCTCCATTCATGCTATGATGACGAGACATC 153755
QY 421 tgcctccagttgacgacgacacccctgggttactggttttagtaactcagtggtgtct 480
|||||
DB 153754 TGCGTCCAGTTGCGACGCTGCATCTGGGTTACTGTTTAGTAACACTGATGCTGTTGCT 153695
QY 481 accctgactctcagctgacactctgtgacccctatgtatagatgaactttctgtgaa 540
|||||

```



Db 153694 ACCCTGACTCTCCAGCTGCACTCTGTGACCCCTATGTGATGATCATTCTCTGTGAA 153635  
 QY 541 gtccctgacgtcctaaattatctgtgttgagacaagaagaagcgaactatc 600  
 Db 153634 GTCCCTGACAGCTCAAGTTACTGTTGTGAGACACAGCAAAAGAGCCTCAACTATTC 153575  
 QY 601 ctgtcagtgagctctccatcaataacccccgacacccatccatatacatgcttt 660  
 Db 153574 CTGTGACATGAGCTCTTCATCATATACCCCTGACACATCATCTTATCATATGCTTTT 153515  
 QY 661 attgtccagagagatltgagatacagtcgtcgtgaaggtcgacaaagacatttgga 720  
 Db 153514 ATTGTCCAGCAGATTTGAGATACAGTCTGTAAGTCGACAAAGACATTTGGGACA 153455  
 QY 721 tgtgttcccatctaatgtgtgtgtctttttatagtaagccgtctgtgtactgt 780  
 Db 153454 TGTGTTTCCCATCTAATGTTGTGTCTTTTATGATACGCGTCTGTGTACCTG 153395  
 QY 781 caaccacttcgcccagctcccaagagcaagaagaatgtgtctctctctatggaac 840  
 Db 153394 CAACCACTTCGCCCAAGCTCCAGAGCAAGAAAGATGTTTCTCTCTATGGAATC 153335  
 QY 841 attgacaccatgctgaatcccccttatatacacttaggaacaagaaggtgaagaagc 900  
 Db 153334 ATTGACCCCATGCTGAATCCCTTATATATACACTTAGGAACAAGAGGTAAAGAGG 153275  
 QY 901 tttaaaagttggtgtgcaagagctctcttaatacaagaataa 942  
 Db 153274 TTTAAAGTTGGTTGCAAGAGTCTTCTTAATCAAGAAATAA 153233

RESULT 3  
 HSA302593 1074 bp DNA PRI 06-JAN-2001  
 LOCUS Homo sapiens 6M1-10\*02 gene for olfactory receptor, cell line  
 DEFINITION KR3598.  
 ACCESSION AJ302593  
 VERSION AJ302593.1 GI:12054410  
 KEYWORDS 6M1-10\*02 gene; olfactory receptor.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1074)  
 AUTHORS Ziegler, A., Ehlers, A., Forbes, S. A., Trowsdale, J.,  
 Uchanska-Ziegler, B., Volz, A., Younger, R. and Beck, S.  
 TITLE Polymorphic olfactory receptor genes and HLA loci constitute  
 (in) extended haplotypes  
 JOURNAL (in) Kasahara, M. (Ed.):  
 MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND  
 FUNCTION: 110-130;  
 Springer-Verlag, Tokyo, Japan (2000)  
 REFERENCE 2 (bases 1 to 1074)  
 AUTHORS Ziegler, A., Ehlers, A., Forbes, S. A., Trowsdale, J., Volz, A.,  
 Younger, R. M. and Beck, S.  
 TITLE Polymorphisms in olfactory receptor genes: a cautionary note  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 1074)  
 AUTHORS Younger, R. M., Amado, C., Bethel, G., Ehlers, A., Fischer, Lindahl, K.,  
 Forbes, S. A., Horton, R., Milne, S., Mungall, A. J., Trowsdale, J.,  
 Volz, A., Ziegler, A. and Beck, S.  
 TITLE Characterisation of clustered MHC-linked Olfactory Receptor Genes  
 in Human and Mouse  
 JOURNAL Unpublished  
 REFERENCE 4 (bases 1 to 1074)  
 AUTHORS Ehlers, A., Beck, S., Forbes, S. A., Trowsdale, J., Volz, A., Younger, R.,  
 and Ziegler, A.  
 TITLE MHC-linked olfactory receptor loci exhibit polymorphism and  
 contribute to extended HLA/OR-haplotypes  
 JOURNAL Unpublished  
 REFERENCE 5 (bases 1 to 1074)  
 AUTHORS Ehlers, A.  
 TITLE Direct Submission

JOURNAL Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik,  
 Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,  
 GERMANY  
 FEATURES  
 source Location/Qualifiers  
 1..1074  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /cell\_line="KR3598"  
 1..1074  
 /gene="6M1-10\*02"  
 1..1074  
 /gene="6M1-10\*02"  
 /codon\_start=1  
 /product="olfactory receptor"  
 /protein\_id="CAC20513.1"  
 /db\_xref="GI:12054411"  
 /translation="MNMYNKSVPQEFLLVPSDQPMLEIPFVMEFSLYITFGNLT  
 ILVSHVDKFLHPTMPEFLSNLSLDLCYTSVPMQVNTKRVISYGGCAOLF  
 IFLAGSRECLLLAVMCDPRVALCRPLHYSITMHQRCFOLAASNTSGFSNYLOS  
 TWLKMPLCGHKEVDHFEVPAALCKSCVDPTNEMAEFTSVFLIPVTLILISY  
 AFIVQAVLRISASGRKAFGCGSHLIVSLFYGTALSMYLOPSPSSDRKRWVSL  
 FCGIILAPMLNPLIYLRNK EYKEAFKRLVAKSLINOETIRNMQISFAKDVLFYLVNF  
 SASCPDIFVTIENYCNLPQRKP"  
 BASE COUNT 277 a 248 c 203 g 346 t  
 ORIGIN  
 Query Match 77.3%; Score 727.8; DB 9; Length 1074;  
 Best Local Similarity 86.4%; Pred. No. 9.7e-198;  
 Matches 815; Conservative 0; Mismatches 127; Indels 1; Gaps 1;  
 QY 1 atgaattgggtaaatagacagcatcacagaagattatctgctggtgtctcagatcga 60  
 Db 1 ATGAATTGGGTAAATAGAGTGTCACAGAGATTCATCTGTAGTTTCTACAGATCAA 60  
 QY 61 ccttgctggagttccacctctctgtgtgtctcttgattctcttaacgtgtgacatctt 120  
 Db 61 CCATGCGTAGAGATTCCACCCCTTGTGTGATGTTGTTCCATATCTTGACAAATCTTT 120  
 QY 121 ggaacatgacattatctctagtgacagcgcctgacacaaactcacaaccccatgtat 180  
 Db 121 GGCATCTGACATATATTTCTTGTCACATGTGGATTTCAATCCACACCCCTATGTATC 180  
 QY 181 ttttctccaatctacactcctcgtgactctgttcttaccacacatgtaacagtcacaa 240  
 Db 181 TTTTCTTAGAATCTCTCACTCGACCTTGCTGTATACCAACAAGTACAGTTCCACAA 240  
 QY 241 atgctagtaaatattagcagcatcaggaagaatcagttacgtgtacgtgtgtaagccag 300  
 Db 241 ATGCTGTAATACATATGACACACAGGAAGAATACATGATGAGTGTGTGCGCCGACG 300  
 QY 301 ctttcataatctctgacctgggggctactgtaataatctctcctgacctgaatgctctt 360  
 Db 301 CTTTTCATTTTCTCTGGCCTTGCGGTTCCACAGATGTCTCTCTGCGCTCATGTGCTTT 360  
 QY 361 gataagttgtagctattgttcgacctcctacactacactcaatgataatgacacagaactc 420  
 Db 361 GATAGATTGTGATATTGTGGGCTTCATCAATTAATATGATGACACAGAGGCTC 420  
 QY 421 tgcctccagttggcagcgacatccctgggttactggttttagtaactcagtggtgtgtct 480  
 Db 421 TGCTTCCAGTTGGCAGCTCATCTCGAATTAATGCTTTAGCAATTAATTAACATGTC 480  
 QY 481 accctgacctccagctcagctcgtgtgacccctatgtatgatatcaatctctctgtga 540  
 Db 481 ACTGTGACACTTAAAGTGCACATGTGTGTGTCACAAAGATGTGATCTTCTTGTGTGA 540  
 QY 541 gtccctgacgtcctaaattatctgtgttgagacaagaagaagaagcgaactatc 600  
 Db 541 GTCCCTGACAGCTCAAGTTACTGTTGTGAGACACAGCAAAAGAGCCTCAACTATTC 600  
 QY 601 ctgtcagtgagctctctcaataacccccgacacccatccatatacatgcttt 660

[illegible]

QY 721 tgggtcccaactaattggtgtgtctcttttataagtaacagccgtctgtgtacctg 780  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 721 TGTGGCTCCCAATCAATATGTGTGTGTCACCTTTTATATGTTACAGTATCCATGACCTG 780  
QY 781 caaccacctgcgcagctcccaaggaacaaagatggttctctctctatgaaac 840  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 781 CAACCACTTCCACCCAGCTCCCAAGACCGGGAAAGATGTTCTCTCTGTGTGATC 840  
QY 841 attaccaccatgcgtgaatcccttatatactactaggaacaaagaaagtaagaaagc 900  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 841 ATTGCACCATGCTGAATCCCTTATATATACCTTAGAACCAAGAGTAAGGAAGCC 900  
QY 901 tttaaaagttgtgtgc-aagagctctcttaatacaagaataa 942  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 901 TTTAAAGTTGTGTGCAAGAGCTTCTTATATCAAGAAATTA 943  
RESULT 5  
HSA302585 1074 bp DNA PRI 06-JAN-2001  
LOCUS Homo sapiens 6M1-10\*01 gene for olfactory receptor, cell line  
DEFINITION BM19.7.  
ACCESSION AJ302585 GI:12054394  
VERSION 6M1-10\*01 gene; olfactory receptor.  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1074)  
Ziegler, A., Ehlers, A., Forbes, S. A., Trowsdale, J., Volz, A.,  
Uchanska-Ziegler, B., Volz, A., Younger, R. and Beck, S.  
TITLE Polymorphic olfactory receptor genes and HLA loci constitute  
extended haplotypes  
(1n) Kasahara, M. (Ed.);  
JOURNAL MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND  
FUNCTION: 110-130;  
REFERENCE 2 (bases 1 to 1074)  
Springer-Verlag, Tokyo, Japan (2000)  
AUTHORS Ziegler, A., Ehlers, A., Forbes, S. A., Trowsdale, J., Volz, A.,  
Younger, R. M. and Beck, S.  
TITLE Polymorphisms in olfactory receptor genes: a cautionary note  
Unpublished  
3 (bases 1 to 1074)  
Younger, R. M., Amadou, C., Bethel, G., Ehlers, A., Fischer, Lindahl, K.,  
Forbes, S. A., Horton, R., Milne, S., Mungall, A. J., Trowsdale, J.,  
Volz, A., Ziegler, A. and Beck, S.  
TITLE Characterisation of clustered MHC-linked olfactory Receptor Genes  
in Human and Mouse  
Unpublished  
4 (bases 1 to 1074)  
Ehlers, A., Beck, S., Forbes, S. A., Trowsdale, J., Volz, A., Younger, R.  
and Ziegler, A.  
TITLE MHC-linked olfactory receptor loci exhibit polymorphism and  
contribute to extended HLA/OR-haplotypes  
Unpublished  
5 (bases 1 to 1074)  
Ehlers, A.  
JOURNAL Direct Submission  
AUTHORS Submitted (06-NOV-2000) Ehlers, A., Institut fuer Immunogenetik,  
TITLES Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,  
GERMANY  
FEATURES  
Source Location/Qualifiers  
1..1074  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/cell\_line="BM19.7"  
1..1074  
/gene="6M1-10\*01"  
1..1074  
/gene="6M1-10\*01"  
/codon\_start=1  
CDS

/product="Olfactory receptor"  
/protein\_id="CAC20505.1"  
/db\_xref="GI:12054395"  
/translation="MNNVKNKSVPOEFLILVNSDQPMLEPIPFVNFILFSLITLTIGNL  
IIVSHVDPKLHTPMYFPLNSLILDLCTYSTVPMILVNCIRKVISYGCVAOLE  
IFALIGSEELCLVLMPCDFREVALPHYSIINHRCFQOLAASRNISGSNSVLDS  
TWTLMPICGHEVDHFECEPALKISCVPTTINEMLEFISVFLIPVTLILISY  
AFIVQAVLRIOASRGKAREGTGSHLIVSLFVGTAFISMTLOPSPSSKRGKMSVL  
FCGLIAPMLNPLITYTLNKEVKEAFKRLVAKSLNQLRNMQISFAKDIYVLTITNF  
SASCPITVITLTCNLPORKEP"  
BASE COUNT 278 a 248 c 202 g 346 t  
ORIGIN  
Query Match 77.1%; Score 726.2; DB 9; Length 1074;  
Best Local Similarity 86.3%; Pred. No. 2.8e-197;  
Matches 814; Conservative 0; Mismatches 128; Indels 1; Gaps 1;  
QY 1 atgaattggaataagacacatcatagagagattatctgtgtgttcagatcga 60  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 1 ATGAATGGGTAATAATAGAGTGTCCACAGAGGTTCATTTGTGATTTCTCAGATCA 60  
QY 61 ccttgctgagattccactccctgtgtgtctctctgattctctacacgtgacacttt 120  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 61 CCATGCTAGAGATTTCCACCCCTTTGATGTTTCTGTTTCTCTATATCTTGCAATCTTT 120  
QY 121 ggaactctgacattatctagtgacgcctgtgacacccaactcaaccatgatat 180  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 121 GGCATCTGACAAATTAATCTTGTGTGCATATGCGATTTCAACCTCACACCCTTATGTAC 180  
QY 181 tttcttaccacatctactcctctgtgattcttctacacacatgtaagctccaca 240  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 181 TTTTTCCTTACCAATCTCTCACTCTGACCTTCTCTCTATACCAAGTACATTCACCA 240  
QY 241 atgcagtaaatatagcagacatcaggaagaatcaatcagatcagtgctgtgtaagccag 300  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 241 ATGCTGGTAATAATATGACACACACAGGAAGTAATACAGTTATGAGGCTGTGGCCAG 300  
QY 301 ctttcataattctgtgcttgggggtactgaataatcttccctgtgcgtatgtccctt 360  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 301 CTTTTCATTTTCCCTGCGCTTGCGGTCACAGATGCTTTCCTGCGGCTCATGTGCTTT 360  
QY 361 gataagttgtatgactattgtgcgcctctccatctcaagtatcatcaccagagac 420  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 361 GATAGCTTTTATGCTATTTTGTGGCTCTCTCACTTCTCATATATATATGACACAGAGCTC 420  
QY 421 tgcctccagttgagccgcgacatccctgggttactagttttagtaactcagtggtgtct 480  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 421 TGTCTCCAGTTGGCAGCTGCAGTCGATCGATTTAGTGGCTTTAGCAATTCAGATTACAGTCC 480  
QY 481 accctgactctcagcgtccactctgtgaacccctatgtgatagtaactctctctgtgaa 540  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 481 ACCTGACACTTAAGATCCACTGTGTGTGTCACAAAGAAAGTGAACACTTCTTCTGTGAA 540  
QY 541 gtccttgacgtctaaatgattatctgtgtgtgagaaagaagaagagagctaaactatct 600  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 541 GTCCCTGCTGTCTCAAGTGTCTGTGTGACACACAGCAAAAGAGGCTAACTATTTC 600  
QY 601 ctgtcagtgagctctcatalcatalaccctcgcgaactcatcatalcatatgattctt 660  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 601 TTCAATCAGTGTCTATCTCTTATATACCGGACACATCATCTTATATCGATGCTTTT 660  
QY 661 attgtcgaagcagatctgaagatagatcgtctgaagtgagcaaaaagcatttgggaca 720  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 661 ATTGTCCAAAGCAGTCTTGAGATTCACAGCTGTGAGAGTCAAGCAAGGCAATTTGGGACA 720  
QY 721 tgggttcccaactaattgtgtgtctcttttataagtaacagccgtctgtgtacctg 780  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 721 TGTGGCTCCCAATCAATATGTGTGTGTCACCTTTTATATGTTACAGTATCCATGACCTG 780  
QY 781 caaccacctgcgcagctcccaaggaacaaagatggttctctctctatgaaac 840  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 781 CAACCACTTCCACCCAGCTCCCAAGACCGGGAAAGATGTTCTCTCTGTGTGATC 840

QY 841 atgcacccatcgtgaatcccttatatacatcacttaggaacaaggaggtlaaagaagc 900  
|||||  
DB 841 ATTGCACCCATGCTGAATCCCTTATATATACACTTAGAGCAAGAGTAAGAAGACC 900  
QY 901 tttaaaggttggttcg-aagagctcttctaatacaagaata 942  
|||||  
DB 901 TTTAAAGGTGTTGCAAGAGCTCTTATCAAGAAATA 943

RESULT 6  
LOCUS HSA302586 1074 bp DNA PRI 06-JAN-2001  
DEFINITION Homo sapiens 6M1-10\*01 gene for olfactory receptor, cell line LG2.  
ACCESSION AJ302586  
VERSION AJ302586.1 GI:12054396  
KEYWORDS 6M1-10\*01 gene; olfactory receptor.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 1074)  
Ziegler, A., Ehlers, A., Forbes, S. A., Trowsdale, J.,  
Uchanska-Ziegler, B., Volz, A., Younger, R. and Beck, S.  
Polymorphic olfactory receptor genes and HLA loci constitute  
extended haplotypes  
(in) Kasahara, M. (Ed.);  
MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND  
FUNCTION: 110-130;  
Springer-Verlag, Tokyo, Japan (2000)  
2 (bases 1 to 1074)  
Ziegler, A., Ehlers, A., Forbes, S. A., Trowsdale, J., Volz, A.,  
Younger, R. M. and Beck, S.  
Polymorphisms in olfactory receptor genes: a cautionary note  
Unpublished  
3 (bases 1 to 1074)  
Younger, R. M., Amadou, C., Bethel, G., Ehlers, A., Fischer, L.,  
Forbes, S. A., Horton, R., Milne, S., Mungall, A. J., Trowsdale, J.,  
Volz, A., Ziegler, A. and Beck, S.  
Characterisation of clustered MHC-linked olfactory Receptor Genes  
in Human and Mouse  
Unpublished  
4 (bases 1 to 1074)  
Ehlers, A., Beck, S., Forbes, S. A., Trowsdale, J., Volz, A., Younger, R.  
and Ziegler, A.  
MHC-linked olfactory receptor loci exhibit polymorphism and  
contribute to extended HLA/OR-haplotypes  
Unpublished  
5 (bases 1 to 1074)  
Ehlers, A.  
Direct Submission  
Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik,  
Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,  
GERMANY

FEATURES  
Source  
1..1074  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/cell\_line="LG2"  
1..1074  
/gene="6M1-10\*01"  
1..1074  
/gene="6M1-10\*01"  
/codon\_start=1  
/product="olfactory receptor"  
/protein\_id="CAC20506.1"  
/db\_xref="GI:12054397"  
/translation="MNWYKSVPOEFILVSDPMLIPEFVLELSYITINGNLT  
ILVSHVDEKLTPTMPEFLSNLSLDLCYTSYVPMVNTCNTRKVISGCVAGOLF  
IFLAGSTECILLAVKCFDRVAICRPLHYSITIMHQRCRFLAASWISGNSVLOS  
TWLKLMPICGHEVDHFECEVPAALKLSCVDTTANEMLEFISVLFILIPVILILISY  
AFIVAVLRISDSAGORAKFGTCSHLIVSLFYGTAIMVLOPSPSSSKDRGNVSL

Query Match 77.1%; Score 726.2; DB 9; Length 1074;  
Best Local Similarity 86.3%; Pred. No. 2.8e-197;  
Matches 814; Conservative 0; Mismatches 128; Indels 1; Gaps 1;  
ORIGIN  
ECGIAPMLNBLIYTLNKEVKEAFKRLVAKSLNNOEIRNNOMISFAKTVLTYLTNF  
SASCPFEVITTEYCNLPORKEP"  
BASE COUNT 278 a 248 c 202 g 346 t  
QY 1 atgaattggttaagacagcatcacacagaggttatctcgtggttcccaatgca 60  
|||||  
DB 1 ATGAATGGGTAAATAAGAGTGTCCACAGAGGATTCATTGTTAGTTTCCACATCA 60  
QY 61 ccttgagctgagttcccaactccttggtctcttgatattcttaacacgtgacatc 120  
|||||  
DB 61 CCATGGCTGAGATTCACACCCCTTGTGAGATGTTCTGTTTCTTATATCTTGCAATCTTT 120  
QY 121 ggcacatctgacattatcttagtgcacgcttgagacacccaattacatcccatg 180  
|||||  
DB 121 GCAATCTGACATTAATTTCTGTGCACATGTGGAATTCACATCACCACCCCTATGTAC 180  
QY 181 tttttcttaacatcatcatcctcgtgactcttggttaacacatgtaagttccaca 240  
|||||  
DB 181 TTTTCTTAAAGCAATCTCTACCTCCGACCTTGTCTATACCAAGTACAGTCCACAA 240  
QY 241 atgcagtaaatcttagcagcatcaggaagaatcaagttacgttcgtggtctgagcc 300  
|||||  
DB 241 ATGCTGTGTAACATATATGACACACCGAAGAAATATAGTTATGTGTGCTGTGTGCCAG 300  
QY 301 ctcttcatactcttgagccttgaggagctactgatactctccttgccgtcatctct 360  
|||||  
DB 301 CTTTCATTTTCCCTGTGGCCCTGTGGTTCCACAGAAATGCTTCTGCGGCTATGTGCTTT 360  
QY 361 gataagttgtgacatctgtcgccctcctcaatcctcaagttatcaatgacacagag 420  
|||||  
DB 361 GATAGGTTGTAGCATTTGTGTGGCCCTCCCTCACTCATATATCATGACACGAGGCTC 420  
QY 421 tgcctcagctgagcagccatccctggtgtactggttttagaacctcagtggtgtct 480  
|||||  
DB 421 TCGTTCACATTTGGACGCTGACATCTGAGATTAATGCTTTAGCAATTCATATTAACATTC 480  
QY 481 accctgacatcctcagctcagcactcgtgaccccatctgtagatagacactctcgtgaa 540  
|||||  
DB 481 ACCGTGACACTTAAGATGACATGTGTGTGACAAAGATGATGATGATGATGATGATG 540  
QY 541 gtccctgacgtcctcaagtatctctgtgttgagacacagcaaatgaggtgacatctc 600  
|||||  
DB 541 GTCCCTGCTCTCTCAAGATGTGCTGTGTGACACAGCAAAATGAGGCTGTAACATATTC 600  
QY 601 ctgtcagtgagctcttcacatcaatcacccctgagacatcatcttatcatatgcttt 660  
|||||  
DB 601 TTTCATCAGTGTCTTATCTTCTTATATACCGTGACATCTCATATCATATCATATCTTT 660  
QY 661 attgctcagcagatctgagatctagcagctcgtgaggtcgacaaagacatcttgagaa 720  
|||||  
DB 661 ATTGTCCAAGCAGTGTGTGAGATTCACAGTGTGTAAGGTCAAGCAAGGATTTGGACA 720  
QY 721 tgtgttcccatcctaattgtgtgtctctttttatagtaacgctcgtctgtgtacctg 780  
|||||  
DB 721 TGTGTCTCCCATCTAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 780  
QY 781 caacacatctgagcagctcagagctcagagcaagaagaatggttctctcttcataagatc 840  
|||||  
DB 781 CAACACCTTCCACCCAGCTCCCAAGACCGGGAAAGATGTTCTCTCTGTGTGAATC 840  
QY 841 attgacaccatcgtgaatcccttatatacatcacttaggaacaaggaggtlaaagaagc 900  
|||||  
DB 841 ATTGCACCCATGCTGAATCCCTTATATATACACTTAGAGCAAGAGTAAGAAGACC 900  
QY 901 tttaaaggttggttcg-aagagctcttctaatacaagaata 942  
|||||  
DB 901 TTTAAAGGTGTTGCAAGAGCTCTTATCAAGAAATA 943

RESULT	7
LOCUS	HSA302587
DEFINITION	HSA302587 1074 bp DNA PRI 06-JAN-2001 Homo sapiens 6M1-10*01 gene for olfactory receptor, cell line H2LCL.
ACCESSION	AJ302587
VERSION	AJ302587.1 GI:12054398
KEYWORDS	6M1-10*01 gene; olfactory receptor. human.
ORGANISM	Homo sapiens
REFERENCE	Eumaxylopt, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 1074)
TITLE	Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J., Uchanska-Ziegler,B., Volz,A., Younger,R. and Beck.S. Polymorphic olfactory receptor genes and HLA loci constitute extended haplotypes (1) Kasahara,M.(Ed.); MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND FUNCTION: 110-130; Springer-Verlag, Tokyo, Japan (2000) 2 (bases 1 to 1074)
JOURNAL	Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J., Volz,A., Younger,R.M. and Beck,S. Polymorphisms in olfactory receptor genes: a cautionary note Unpublished 3 (bases 1 to 1074)
REFERENCE	Younger,R.M., Amadou,C., Bethel,G., Ehlers,A., Fischer Lindahl,K., Forbes,S.A., Horton,R., Milne,S., Mungall,A.J., Trowsdale,J., Volz,A., Ziegler,A. and Beck,S. Characterisation of clustered MHC-linked Olfactory Receptor Genes in Human and Mouse Unpublished 4 (bases 1 to 1074)
TITLE	Ehlers,A., Beck,S., Forbes,S.A., Trowsdale,J., Volz,A., Younger,R. MHC-linked olfactory receptor loci exhibit polymorphism and contribute to extended HLA/OR-haplotypes Unpublished 5 (bases 1 to 1074)
JOURNAL	Ehlers,A. Direct Submission Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik, Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin, GERMANY
FEATURES	Location/Oualifiers source 1..1074 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="6" /cell_line="H2LCL" 1..1074 /gene="6M1-10*01" 1..1074 /gene="6M1-10*01" /codon_start=1 /product="olfactory receptor" /protein_id="CAC20507.1" /db_xref="GI:12054398" /translation="MMWYKNSVPDEFLILVFSDDPMLEIPFVFELSYILITFGNLTLL IIVSHDFEKLHTPMVEFLSNLSLDLCYSTSTPOMLVNCTNRKYISGCVAAOLE IFLAGSEECILLAVMGCDPRVALICPLRLYSIIHNRICPOLAASMTSGSNSLOS TWTKLMPLCGKHREVDHFPCPEPALIKSCVDVTGNARELFPTSVLPILLYPTTLLISTY AFIVQAVLRIDSSAGORAKFCFTGGSHLIWSLFTGIALSMLOPPSPSSKRGKMWISI FCGIIAPMLNPILTYTLRNKEVEKAFFKRLVAKSLLINOETIRNMQMSIFAQDVLTYLTNE SASCPIPVLTIEENCNLPORKFP"
CDS	278 a 248 c 202 g 346 t
BASE COUNT	
ORIGIN	
Query Match	77.1%; Score 72.2; DB 9; Length 1074;

	Best Local Similarity	86.3%	Pred. No. 2.8e-197;	
	Matches	814; Conservative	0; Mismatches	128; Indels
QY	1	atgaattggtgaaatgacagcatalcatagaggttattctgtcgggttctcgaatcga	60	
Db	1	ATGAAATGGGTAATAATGAAGAGTGTCCACAGAGAGTTCAATCTGTAGTTTCTCAGATCAA	60	
QY	61	ctgtgagcgaggttccactcctgtgtgtcttcttgatttttctaacgtgacatcttt	120	
Db	61	CCATGGCTAGAGATTCACCCCTTGTGTAGATTTCTGTTCCTTAATCTTAATCAATCTTT	120	
QY	121	ggcaatcgtaccattatctagtgtcaagcctgtgacacacaaactctacccccaatgtat	180	
Db	121	GGCAATCTGACAAATATTTCTGTGTGCATATGTGATTTCAAACTGCCAACCCCTATATGAC	180	
QY	181	tttttcttacaactatcatcactcctgtgacttctgttacaacaaatgtacagtcacca	240	
Db	181	TTTTTTCTTACCAATCTCTCACTCCTGAGACCTTTGGCTATACCAAGATCAGTTCCACAA	240	
QY	241	atgtagtagaaattatgcaagcatcaagaaatcaatcagttactgtgtgtgtgtgcccag	300	
Db	241	ATGTGTGTAATAACATATATGACACACAGGAATATCAATTAATGTGTGTGTGTGGCCAG	300	
QY	301	ctttcattatcttcggcctctgtggggtacgtgaatactctcccgccgcatgtcctt	360	
Db	301	CTTTTCATTTTCTGTGGCTTGTGGTTCCACAAATGTCTTCTCTGGCGGTATGTGCTTT	360	
QY	361	gataagtttgtagctatttctgcgcctctccattactcaattcaatcaltgacacaaagac	420	
Db	361	GATAGGTTTGTAGCATTTGTGTGGCCTCTCCATTACTCAATTAATCATGACACAGAGGCTC	420	
QY	421	tgtcctcagttgtgcagccgcacactcctgtgttactcgtgttttaactcagttgtgtct	480	
Db	421	TGCTTCCAGTTGTGGAGCTGCATCCTGTGATTTGTGTGGCTTTTACCAATCTGATTTACAGTCC	480	
QY	481	accctgacctcccaagctgcacactcgtgaaccctatgtgatalcatcactctcgtgaa	540	
Db	481	ACCTGGACACTTAAGATGCCACTGTGTGTGTCACAAAGAAGTGGATCACTTCTTCTGTAA	540	
QY	541	gtccctgacctgtcctaagttactctgtgtgttagaacaacagaanaatgagctgaactatc	600	
Db	541	GTCCCTGCTGTGTCACAAATGTCCTGTGTGACACAAAGCAAGCAATGAGCGTGAACATAATTC	600	
QY	601	cttgtcagtgagctcttccatctaatcaccctgacacatcactatcatatagctttt	660	
Db	601	TTTCATCACTGTGTCTATCTCTTAATACCCGTGACACTCACTCTTAATGTCTGATCTCTTT	660	
QY	661	atttctcgagcaagatattagatatacagctcgtctaagttcgaaacaaagcatttgggaca	720	
Db	661	ATTGTCCAGAGAGTGTGAATTCAGTCTGTGAAGTGCACAAAGCAAGCATTTGGGACA	720	
QY	721	tgtgtgtcccatctaatctgtgtgtctcttlltaatgataagccggtctcgtgtactg	780	
Db	721	TGTGTGCTTCCATCTAATTTGTGTGTGCTCACTTTTATATGTTACAGATATCTCCATGTACTG	780	
QY	781	caacacactcgcagcactcaaggaaccaagaaatgtgttctctctctatagaaatc	840	
Db	781	CAACACCTTCAACAGCTCCAAAGACCGGGGAAAGATGTCTCTCTCTGTGTGATTC	840	
QY	841	attgacaccaatgctgaaatcccttatatatatacacttaagaaacaaggaagttaagaaagc	900	
Db	841	ATTGACACCAATGTGAATCCCTTAATAATATACCTATTAGAACAAAGAGGTAAAGGAACCC	900	
QY	901	tttaaaaggttgatgtgc--aagatctctcttaatcaagaataaa	942	
Db	901	TTTAAAGGCTTGTGTGCAAGAGTCTTTTAATCAAGAAATRA	943	

RESULT

8

LOCUS

HSA302588

DEFINITION

Homo sapiens 6M1-10\*01 gene for olfactory receptor, cell line WT51.

ACCESSION

AJ302588



Uchanska-Ziegler, B., Volz, A., Younger, R. and Beck, S.  
 Polymorphic olfactory receptor genes and HLA loci constitute  
 extended haplotypes  
 (in) Kasahara, M. (Ed.):  
 MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND  
 FUNCTION: 110-130;  
 Springer-Verlag, Tokyo, Japan (2000)

REFERENCE  
 AUTHORS  
 Ziegler, A., Ehlers, A., Forbes, S. A., Trowsdale, J., Volz, A.,  
 Younger, R. M. and Beck, S.  
 Polymorphisms in olfactory receptor genes: a cautionary note  
 3 (bases 1 to 1074)  
 Younger, R. M., Amadou, C., Bethel, G., Ehlers, A., Fischer, Lindahl, K.,  
 Volz, A., Ziegler, A. and Beck, S.  
 Characterisation of clustered MHC-linked Olfactory Receptor Genes  
 in Human and Mouse  
 Unpublished  
 4 (bases 1 to 1074)  
 Ehlers, A., Beck, S., Forbes, S. A., Trowsdale, J., Volz, A., Younger, R.  
 and Ziegler, A.  
 MHC-linked olfactory receptor loci exhibit polymorphism and  
 contribute to extended HLA/OR-haplotypes  
 5 (bases 1 to 1074)  
 Ehlers, A.  
 Direct Submission  
 Submitted (06-NOV-2000) Ehlers, A., Institut fuer Immunogenetik,  
 Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,  
 GERMANY

FEATURES  
 source  
 1..1074  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /cell\_line="SA"  
 1..1074  
 /gene="6M1-10\*01"  
 1..1074  
 /gene="6M1-10\*01"  
 /codon\_start=1  
 /product="olfactory receptor"  
 /protein\_id="CAC20509.1"  
 /db\_xref="GI:12054403"  
 /translation="MMWVNSVPOEFILVSDPWEIPPEVLEPSYILITFGNLT  
 IIVSHVDFKLTMYFSLNSLDLCYTSIPOMLVNCFNKRKYSYGCVAOLF  
 IFIALGSTECILLAVMCEDFVAICRPLHYSIIHMRCLPOLAASMTSGSNVLOS  
 TWLTKMPLCGHKEVDHPCCEPALIKSLCVDPTANEAELFISVFLIPVLTILISY  
 AFIVQAVLRIDSAEGORAKRGTCGSHLIVSLFYGTALSMILOPPSSKRGKAVSL  
 FCGIIAPMLNPILITLRNKEVKAFLRLVANSLLNOETIRNMOMISFADTVLTVTNF  
 SASCPITVITLNNCNLPQRKFP"

CDS  
 gene  
 1..1074  
 /gene="6M1-10\*01"  
 1..1074  
 /gene="6M1-10\*01"  
 /codon\_start=1  
 /product="olfactory receptor"  
 /protein\_id="CAC20509.1"  
 /db\_xref="GI:12054403"

BASE COUNT 278 a 248 c 202 g 346 t  
 ORIGIN

Query Match 77.1%; Score 126.2; DB 9; Length 1074;  
 Best Local Similarity 86.3%; Pred. No. 2.8e-197;  
 Matches 814; Conservative 0; Mismatches 128; Indels 1; Gaps 1;

QY 1 atgaatggtaaatagacagcatcatagaggttatctctggtggtttcccaatcga 60  
 DB 1 ATGAATGGTAATAAGAGTGTCCACAGGAGTTCATTCTGTGTTTTCACAGATCA 60  
 QY 61 ccttgagcgaatccctctgtgtctctctgattcttacaactggaacatctt 120  
 DB 61 CCATGGCTAGAGATTCCACACCTTTGTGATGTTCTGTTTCTTCTAATCTTGACATCTTT 120  
 QY 121 ggaacatcgacattatctagtgtaacgcctcgagacacaaactcatacccatglat 180  
 DB 121 GGCAATCGACAAATATCTTGTGTACATGTGATTTCAAACTCCACACCCCTATGTAC 180  
 QY 181 ttcttctacacattcatctctgtgattctgtttacacacatgttccacaa 240

|||||  
 DB 181 TTTTCTTAGAATCTCTCACTCCTCGACCTTTCCTATACCAAGATGACATTCACAA 240  
 QY 241 atcgtagaataatitcgaacalcaggaagaatcagtlacgtgctgtgtagccag 300  
 DB 241 ATGCTGTAAACATATGACACACACAGGAAGAAATCAGTATAGTGCGCTGTGCGCCAG 300  
 QY 301 ctttcataattctgacctgggggctactgataattctctccgctgcatgctc 360  
 DB 301 CTTTTCATTTTCTCTGCGCTGTGCGCTCAGAGATATCTTCTCTGCGCTCATGTGCTTT 360  
 QY 361 gataagtttgaatctatctgtgacctccatctcactcaagtlacatcagcagaagctc 420  
 DB 361 GATAGTTTGTATGCTATTTGTGCGCTCTCCATTTCTCATATATGATGACAGAGGCTC 420  
 QY 421 tgcctcagttggaacgcacactcgtgggttactggttttagtaactcagtggtgtct 480  
 DB 421 TGCTTCCAGTTGGCAGCTGCATCGATAGTGTAGTGCCTTACGAAATTCAGTATTCAGTCC 480  
 QY 481 accctgacctccacgtcgcacctctgtgacccctatgtagatgacattctctgtgaa 540  
 DB 481 ACCCTGACACTTAAGATGCCACTGTGTGTCAAAAGAGTGAATCATTCTTGTGAA 540  
 QY 541 gtcctcagctgctcaagtlacctgtgtgtagacacagcaacagagctgaactatctc 600  
 DB 541 GTCCCTGCTCTCTCAAGTGTGTCTGTGTGACACAAAGCAATGAGGCTGAATCTTC 600  
 QY 601 ctgtcagtgagctctctccatctaatccctgacactcattatcatatgctctt 660  
 DB 601 TTCATCAGTGTCTATTCCTCTTAATACCCGTCGATCCTTATATGATGATCTT 660  
 QY 661 attgctcagagcagatltgagatcacgtctgtgaaagtcgacaaaagcatltgggaa 720  
 DB 661 ATTGTCACAGCAGTGTGGAATCCAGTCTGTGAAGTCAACGAAAGCATTTGGGACA 720  
 QY 721 tgtgttcccatctaatgtgtgtctctctttagtagacagcgctcgtgtgacctg 780  
 DB 721 TGTGCTCCCATCTAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 780  
 QY 781 caacacacttcgcccagctcgaagcagaaggaagaatgttctctctctctatgaaac 840  
 DB 781 CAACACACTTTCACAGCTCCAAAGACGGGAAAGATGTTTCTCTCTGTGTGAATC 840  
 QY 841 attgacacatgctgaatcccttatatacatcattgagaacaaagaggttaagaagcc 900  
 DB 841 ATTGACACCATCTCAATCCCTTATATATACACTTAGCAACAAAGAGTAAAGAGGCC 900  
 QY 901 ttttaaaggttggttgc-aagagctctcttaataagaanaaa 942  
 DB 901 TTTTAAAGGTGTGTGCAAGAGTCTTCTTATACAGAAATATA 943

RESULT 10  
 HSA302590 1074 bp DNA PRI 06-JAN-2001  
 LOCUS HSA302590  
 DEFINITION Homo sapiens 6M1-10\*01 gene for olfactory receptor, cell line YAR.  
 ACCESSION AJ302590  
 VERSION AJ302590.1 GI:12054404  
 KEYWORDS 6M1-10\*01 gene; olfactory receptor.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 1074)  
 Ziegler, A., Ehlers, A., Forbes, S. A., Trowsdale, J.,  
 Uchanska-Ziegler, B., Volz, A., Younger, R. and Beck, S.  
 Polymorphic olfactory receptor genes and HLA loci constitute  
 extended haplotypes  
 (in) Kasahara, M. (Ed.):  
 MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND  
 FUNCTION: 110-130;  
 Springer-Verlag, Tokyo, Japan (2000)  
 2 (bases 1 to 1074)

AUTHORS Ziegler, A., Ehlers, A., Forbes, S.A., Trowsdale, J., Volz, A.,  
 Younger, R.M., and Beck, S.  
 TITLE Polymorphisms in olfactory receptor genes: a cautionary note  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 1074)  
 AUTHORS Younger, R.M., Amdou, C., Bethel, G., Ehlers, A., Fischer Lindahl, K.,  
 Forbes, S.A., Horton, R., Milne, S., Mungall, A.J., Trowsdale, J.,  
 Volz, A., Ziegler, A., and Beck, S.  
 TITLE Characterisation of clustered MHC-linked Olfactory Receptor Genes  
 in Human and Mouse  
 JOURNAL Unpublished  
 REFERENCE 4 (bases 1 to 1074)  
 AUTHORS Ehlers, A., Beck, S., Forbes, S.A., Trowsdale, J., Volz, A., Younger, R.  
 and Ziegler, A.  
 TITLE MHC-linked olfactory receptor loci exhibit polymorphism and  
 contribute to extended HLA/OR-haplotypes  
 JOURNAL Unpublished  
 REFERENCE 5 (bases 1 to 1074)  
 AUTHORS Ehlers, A.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-NOV-2000) Ehlers, A., Institut fuer Immunogenetik,  
 Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,  
 GERMANY

FEATURES  
 source Location/Qualifiers  
 1..1074  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /cell\_line="YAR"  
 1..1074  
 /gene="6M1-10\*01"  
 1..1074  
 /gene="6M1-10\*01"  
 /codon\_start=1  
 /product="olfactory receptor"  
 /protein\_id="CAC20510.1"  
 /db\_xref="GI:12054405"  
 /translation="MNNVNSVPOEFLIVSDPFLIEPFLVFLSYITIEFGLT  
 IIVSHVDFKLTMPYFLSNLSLDCTTSTVOMLVNICTNTKVSISGCAVOLP  
 IFIALGSTECILIAWCFDAVATCRPHYSITIMORICPOLAASMSISGNSVLOS  
 TWTLKMPICGKHVDHFECEVPALEKLSVDTAEALFETISVFLIPYLLISTY  
 APTVAVLRISASQSKRAFGICGSHLIVSLFTGTAISMTLQPSPSKDKQWVSL  
 FCGIAPLNPILITLRNKEVKAERLVAKSLNQLRNMOMISFADTVLYLTNF  
 SASCPITFTEENYCNLPORKEP"  
 BASE COUNT 278 a 248 c 202 g 346 t  
 ORIGIN

Query Match 77.1% Score 726.2, DB 9; Length 1074;  
 Best Local Similarity 86.3% Pred. No. 2,8e-197;  
 Matches 814; Conservative 0; Mismatches 128; Indels 1; Gaps 1;

QY 1 atgaattggtaagacgacatcacacaggttatctctggtttccagatcga 60  
 |||||||  
 DB 1 ATGAATGGGTAATGAAGGTGTCCACAGGAGTCAATCTGTAGTTTTCAGATCAA 60  
 |||||||  
 QY 61 ccttggtgaggttccacacctgtgtctcttctgattcttaccagtgaaccattt 120  
 |||||||  
 DB 61 CCATGGGTAGAGATTCCACCCCTTTGTGATGTTCTGTTTCTTATATCTTGACAAATCTTT 120  
 |||||||  
 QY 121 ggcacatcgacatctattctagtcacgctcgtgacacaaacttatacccccattat 180  
 |||||||  
 DB 121 GCCAATCTGACAAATATATTCTGTGCACATGTGATTTCAAACTCCACACCCCTATGTAC 180  
 |||||||  
 QY 181 tttttcttaccacatctcactcctcgtgattgttttaccacacatgtacatccacaa 240  
 |||||||  
 DB 181 TTTTCTTAGCAATCTCTCACTCCCTGGACCTTTGCTATACCAACAAGTACAGTTCCACAA 240  
 |||||||  
 QY 241 atgtagtaaaattatgtagcagcatcaggaagaatcatcagtcagtcgtgttagccag 300  
 |||||||  
 DB 241 ATGCGGTAAACATATGCAACACCAAGAAAGTATCATGATTATGCTGTGTGGCCGAG 300  
 |||||||  
 QY 301 ctttcataattcttggtccttggtggtcctactgaatactctccctggtcgtcatgtc 360  
 |||||||

|||||  
 DB 301 CTTTCATTTCCTGCGCTTGGGTTCCACAGATGCTCTCTGCGCCCTCATGTCCTT 360  
 |||||||  
 QY 361 gatagtttgtatgataattgttggtgctctccattactcaatgataatcagacagac 420  
 |||||||  
 DB 361 GATAGTTGTGTAGCTATTTGTGGCCTCTCCATTTACTCAATATCATGACCAAGAGGCTC 420  
 |||||||  
 QY 421 tgcctcagttgacacgcatccctgggttactaggttttagtaacacagtcgtgtgtct 480  
 |||||||  
 DB 421 TGCTTCACTTTGGCAGCTCTCATCTCGATTAGTGGCTTTAGCAATTTCACTATTACAGTCC 480  
 |||||||  
 QY 481 accctgactctcagctcgtcactctgtgacccctatgtatagatcaattctcgtgaa 540  
 |||||||  
 DB 481 AACTGACACCTTAAGATGCGACACTGTGTGTCACAAAGAGTGATGATCTTCTGTGAA 540  
 |||||||  
 QY 541 gtcctcgtcactgtcactgataattctgtgtgttgagacacagcaaaagaggtgactatc 600  
 |||||||  
 DB 541 GTCCCTGCTCTGCTCAAGTTGTCTGTGTGACACAAACGCAAAATGAGCTCAACTATTC 600  
 |||||||  
 QY 601 ctgtcagtgagctctctcactaataacccctgacacactcctatcatcatgtctt 660  
 |||||||  
 DB 601 TTCAATCAGTGTCTATTCTTCTTAATACCGGTGACACTCATCTTATATGCTATGCTTT 660  
 |||||||  
 QY 661 attgtccgagcaglatltgagatagacagtcgtcgtgaaagtcgacaaagacattgagaca 720  
 |||||||  
 DB 661 ATTGTCCAGCAGTGTGTAATCCAGTCTGCTGAAGTCACAAAGGATTTGGGACA 720  
 |||||||  
 QY 721 tttgtgtccatctaatgtgtgtgtctcttttattagtagacagcgtctcgtgtactgt 780  
 |||||||  
 DB 721 TGTGTCTCCATCTAATTTGTGTGTGTCATTTTATTTATGTCAGCTATTCCTCATGACCTG 780  
 |||||||  
 QY 781 caacacacttgcggcagcctcacaagacacaaagaaatgttctctctcattgaaac 840  
 |||||||  
 DB 781 CAACCACTTTCACCCAGCTCCCAAGACCGGGAAGATGTTTCTCTTCTGTGCAATC 840  
 |||||||  
 QY 841 attgcacccatgctgaatcccttatatactacttaggaacaagaggtlaaagaagtc 900  
 |||||||  
 DB 841 ATTGCACCCATGCTGAATCCCTTATATATACACTTAGGAACAAGAGTAAAGAACCC 900  
 |||||||  
 QY 901 tttaaaagtgtgtgtgc-aagaagctcttcaatcaagaataa 942  
 |||||||  
 DB 901 TTTAAAGTGTGTGCAAAAGAGCTTCTTATCAAGAAATAA 943  
 |||||||

RESULT 11  
 HSA302591  
 LOCUS HSA302591 1074 bp DNA PRI 06-JAN-2001  
 DEFINITION Homo sapiens 6M1-10\*01 gene for olfactory receptor, cell line OLGA.  
 ACCESSION AJ302591  
 VERSION AJ302591.1 GI:12054406  
 KEYWORDS 6M1-10\*01 gene; olfactory receptor.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1074)  
 Ziegler, A., Ehlers, A., Forbes, S.A., Trowsdale, J.,  
 Uchanska-Ziegler, B., Volz, A., Younger, R. and Beck, S.  
 TITLE Polymorphic olfactory receptor genes and HLA loci constitute  
 extended haplotypes  
 (in) Kasahara, M. (Ed.);  
 MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND  
 FUNCTION: 110-130;  
 Springer-Verlag, Tokyo, Japan (2000)  
 REFERENCE 2 (bases 1 to 1074)  
 Ziegler, A., Ehlers, A., Forbes, S.A., Trowsdale, J., Volz, A.,  
 Younger, R.M., and Beck, S.  
 TITLE Polymorphisms in olfactory receptor genes: a cautionary note  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 1074)  
 Younger, R.M., Amdou, C., Bethel, G., Ehlers, A., Fischer Lindahl, K.,  
 Forbes, S.A., Horton, R., Milne, S., Mungall, A.J., Trowsdale, J.,  
 Volz, A., Ziegler, A., and Beck, S.



TITLE Characterisation of clustered MHC-linked Olfactory Receptor Genes in Human and Mouse

JOURNAL Unpublished

REFERENCE 4 (bases 1 to 1074)

AUTHORS Ehlers, A., Beck, S., Forbes, S.A., Trowsdale, J., Volz, A., Younger, R. and Ziegler, A.

TITLE MHC-linked olfactory receptor loci exhibit polymorphism and contribute to extended HLA/OR-haplotypes

JOURNAL Unpublished

AUTHORS Ehlers, A.

REFERENCE Direct Submission

TITLE Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik, Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin, GERMANY

JOURNAL

FEATURES

Source Location/Qualifiers

1..1074

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="6"

/cell\_line="OLGA"

1..1074

/gene="6M1-10\*01"

1..1074

/gene="6M1-10\*01"

/product="Olfactory receptor"

/protein\_id="CAC20511.1"

/db\_xref="GI:12054407"

/translation="MMWVNSVDPDEFLLVSDPMLFPPFVLFSLYLLTFTGNTLLVSHVDFLHTPMYFPLNSLLDLCTYSYVPMVINCRTKYSYGGCAOLFIFLAGSTECILLAVMCFDRFVACRPLHYSIIHQRCLCOLAASMSISGNSVLDS TWLTKMPLGCHKEVDHFCVDFCEPDLKISITANREAFESILYLLIPYLLISYAFIYQAVLRIGSASGOKRAGTOSSHIIVSLPFGTISIMTLOPSPSSSKDRGMVSLFCGIIAPMLNPLVLTLLRNKEKFAFKLVAKSLNDEIRNMOMISFAKDVLYLYTNF SASCPFIVIRIENCNLPORRFP"

BASE COUNT 278 a 248 c 202 g 346 t

ORIGIN

Query Match 77.1%; Score 726.2; DB 9; Length 1074;

- Best Local Similarity 86.3%; Pred. No. 2.8e-197;

Matches 814; Conservative 0; Mismatches 128; Indels 1; Gaps 1;

QY 1 atgaattggtgaatgaagcatcatcagaagattatctgcgtgggtcttcaatcga 60

DB 1 ATGAATTGGTAATGAAGAAGTCCACAGAGATTTCATTGTTTCTTCAGATCAA 60

QY 61 ccttgagctgaggttccactccttctgtcttcttgaattcttcaactgtgacatctt 120

DB 61 CCATGGCTAGAGATTCCACCCTTGTGTGATGTTCTGTTTCTTCTATATCTTGACAATCTTT 120

QY 121 ggcacatcgacatattatagtagcagcctggaacacaaatcatatccccatgta 180

DB 121 GGCAATCGACAATATCTGTGTGACATGATTCATTCACAATCCACACCCCTATGTAC 180

QY 181 tttttcttacaatctatcaactcctgtgatttcttcaacacatgtgaacgtccaca 240

DB 181 TTTTCTTACCAATCTCTCACTCTCTGACCTTGTGCTATACCAACAGATCAAGTTTCCACAA 240

QY 241 atgctagtaaatattatgagcatcaggaagaatcatcagttatcgtggtctgttaaccag 300

DB 241 ATGCTGGTAACATATGACACACCGAAGTAATCATCTATATGCTGTGTGCTGACCCAG 300

QY 301 cttctcatattcttgacctggtgggtactgaataatcttctcctggtgacatgtcctt 360

DB 301 CTTTTCATTTTCTGCTGGCTTGGTCCACAGATGTCTTCCTGCGCTCATGTGCTTT 360

QY 361 gatagatttgaatatttgcgaccttcaattactactatcagttatcatcacaagaatc 420

DB 361 GATGAGTTTGAAGCTATTGTGGCTCTCTCACTTAATCATATATCATCCACAGAGGCTC 420

QY 421 tgcctcagattgagccgcatcctcgtgttactggttttagtaactcagtggtgtct 480

DB 421 TGCTTCCAGTTGGCAGCTGCATCCCGATGATGAGCTTTAGCAATTCAGTATACAGTCC 480

QY 481 accctgactctccagctgcactctgtgaacccctatgtgataagatcacttctctgtaa 540

DB 481 ACCCTGACTCTCCAGCTGCATCCCGATGATGAGCTTTAGCAATTCAGTATACAGTCC 540

QY 541 gtccctgagcagctcaatattatctgtgtgtgagacaagaagaatgagctgaactatc 600

DB 541 GTCCCTGCTTGTCTCAATTTGTCTGTGTGACACACAGCAATGAGGCTGACTATTC 600

QY 601 ctgtcagtgagctcttccatcaatatacccttgacactcatctatatacatatgcttt 660

DB 601 TTCATCAGTGTGATATCTCTCTTATATACCCGACACATCATCTTATATGATGATCTTT 660

QY 661 attgtcagcagatattgagatatacagctctgtgaagtgagcaaaagaatttggaca 720

DB 661 ATTGTCCAGCAGATGTTAGAAATCCAGCTCTGAGGCTCAAGGCAATTTGGGACA 720

QY 721 tgtgttccatcattatgtgtgtgtctcttctttagtagtaagcgtctgtgactg 780

DB 721 TGTGCTCCCATCTTATTTGTGTGTCTCTTTTATGTGTAGCTATCTTCATGTACTG 780

QY 781 caaccaccttgcagcagctcccaaggaagcaagaatgttctctcttctatagaatc 840

DB 781 CAACCACTTCACCCAGCTCCCAAGAGCGGGAAAGATGTTCTCTCTGTGTGATC 840

QY 841 attgacacatgctgaatcccttatatacacttagaacaagaagtaagaagc 900

DB 841 ATTGACCACTTCAGTAATCCCTTATATATACCTTAGGACCAAGAGTAAGGAGAGCC 900

QY 901 tttaaagattgtgtgc-aagagcttcttcaatgaagaataa 942

DB 901 TTTAAAGATTGTGTGCAAGAGCTCTTATATCAAGAAATPA 943

RESULT 12

LOCUS HSA302592 1074 bp DNA 06-JAN-2001

DEFINITION Homo sapiens 6M1-10\*01 gene for olfactory receptor, cell line AMAT.

ACCESSION AJ302592

VERSION AJ302592.1 GI:12054408

KEYWORDS 6M1-10\*01 gene; olfactory receptor.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1074)

AUTHORS Ziegler, A., Ehlers, A., Forbes, S.A., Trowsdale, J., Uchanska-Ziegler, B., Volz, A., Younger, R. and Beck, S.

TITLE Polymorphic olfactory receptor genes and HLA loci constitute extended haplotypes

JOURNAL (in) Kasahara, M. (Ed.); MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND FUNCTION: 110-130; Springer-Verlag, Tokyo, Japan (2000)

REFERENCE 2 (bases 1 to 1074)

AUTHORS Ziegler, A., Ehlers, A., Forbes, S.A., Trowsdale, J., Volz, A., Younger, R.M. and Beck, S.

TITLE Polymorphisms in olfactory receptor genes: a cautionary note

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 1074)

AUTHORS Younger, R.M., Amandou, C., Bethel, G., Ehlers, A., Fischer, Lindahl, K., Forbes, S.A., Horton, R., Milne, S., Mungall, A.J., Trowsdale, J., Volz, A., Ziegler, A. and Beck, S.

TITLE Characterisation of clustered MHC-linked Olfactory Receptor Genes in Human and Mouse

JOURNAL Unpublished

REFERENCE 4 (bases 1 to 1074)

AUTHORS Ehlers, A., Beck, S., Forbes, S.A., Trowsdale, J., Volz, A., Younger, R. and Ziegler, A.

TITLE MHC-linked olfactory receptor loci exhibit polymorphism and contribute to extended HLA/OR-haplotypes

JOURNAL Unpublished  
REFERENCE 5 (bases 1 to 1074)  
AUTHORS Ehlers A.  
TITLE Direct Submission  
JOURNAL Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik,  
Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,  
GERMANY

## FEATURES

## source

Location/Qualifiers  
1..1074  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/cell\_line="AMAI"  
1..1074  
/gene="6M1-10\*01"  
1..1074  
/gene="6M1-10\*01"  
/gene="6M1-10\*01"  
/codon\_start=1  
/product="Olfactory receptor"  
/protein\_id="CAC20512.1"  
/db\_xref="GI:12054409"  
/translation="MNMYNKSVPDFILLVSDPMLIEPFPVNFSLYITITGNLT  
IIIVSHDFKLIHTPMFPLSLSLDLCTTSTVPMLVNTCNKRVISTGGVAOLF  
IFLALGTECHLLAVMCFDRFVAICRPLHSIIMHQRLEQLAASWISFSNSVLOS  
TWLTKMPCGHEVDHFEVPALEKLSVDYTNEMAELEFISVLFILIPYLLISY  
AFIVQAVLRISAGORKAFCGSHLIVLVEFGTAISMLOPPSPSKRMVSL  
FCGIAPMLNPLIYLRNKEVKAERFLVAKSLNQEIRNNQMISPAKDIYLVLTNF  
SASCPFIYITENYCNLPQRKEP"

## gene

## CDS

BASE COUNT 278 a 248 c 202 g 346 t  
ORIGIN

Query Match 77.1% Score 726.2; DB 9; Length 1074;  
Best Local Similarity 86.3% Pred. No. 2,8e-197;

Matches 814; Conservative 0; Mismatches 128; Indels 1; Gaps 1;

OY 1 atgaattggtaaatgaacagcatcacagaagattatctgctgggttcacagatcga 60  
DB 1 ATGAATTGGGTAAATGAAGCTGCCACGAGATTCATCTGTAGTTTCTACATCA 60  
OY 61 ccttgacggagattccactcctctgctctctgattcttctaacgtgacatcttt 120  
DB 61 CCATGGCTAGAGATTCCACCTCTGTGATGTTCTGTTTCTATATCTTGACAAATCTTT 120  
OY 121 ggaacatcgaccattatctagtgcaagcctcgacacacacattcatacccatgtat 180  
DB 121 GGCATCTCGACAAATATCTTGTGTACATGTGATTTCAATCTCACACCCCTATGTAC 180  
OY 181 ttttcttaccatctacactcctctgattctgttaccacacatgtaagtcacaaa 240  
DB 181 TTTTCTTACCAATCTCACTCTGACCTTGACCTTGTCTATACCAAGTACATTCACAAA 240  
OY 241 atgcagtaaatatttgacagcatcgaagaagaatcagttatcgtgctgtgtaacag 300  
DB 241 ATGCTGGTAACATATGCAACACGAGAAATGATGATGAGTGTGTGCTGACG 300  
OY 301 ctttccatattctggccttgggggcttactgaatatcttccctggccgtcatgfcctt 360  
DB 301 CTTTTCATTTTCTGGCTTGGCTTGGCTTCACAGAAATGTCTTCTCTGCGCTCATGTGCTTT 360  
OY 361 gataagtttgatgattctgtgagccttcacattactacagttatcatcagacagagac 420  
DB 361 GATAGATTGATGATTTGTTGGCTCTCTCCATTACTCATTTATTCATGACACAGAGGCTC 420  
OY 421 tgcctcagttggcagcgcgacactcctggttactggttttagtaactcagtggtgtgtc 480  
DB 421 TGCCTCCAGTTGGCAGCTGCATCTCGATGATGAGCTTACCAATTCAGTATTAACAGTCC 480  
OY 481 acccagactccagctcagctcgtgaccccatgtagatagatcattcttcgtgtaa 540  
DB 481 ACCCTGGACACTTAAGATCCACTGTGTGTACAAAGATGAGTCACTTCTTGTGTAA 540  
OY 541 gtccctgactgctcaagttaattctgtgtgagacacagcaaatgaggtcgaactatcc 600

DB 541 GTCCCTGGCTCTGCTCAAGTTGTCGTGTGACACACAGCAAAATGAGGCTCACTATTC 600  
OY 601 ctgtcagttgagctcttcacattacacccctgacacccaacccattatcatatgctttc 660  
DB 601 TTCATCAGTGTGCTATTCCTTCTTAATACCCGAGACACATCATCTTATATCATCTTTT 660  
OY 661 attgtccagcagattttagagatagacagctgtgtaaggtcgcacaaagacatttggaca 720  
DB 661 ATTGTCCAAGCAGTGTGAGAAATCCAGTCTGCTGAAGTCAACGAAGACATTTGGACA 720  
OY 721 tgtgtctccacatcaattgtgtgtctcttcttctttagtaacagcgtctctgtgacactg 780  
DB 721 TGTGCTCCCATCTAATGTGTGTGTGTCACATTTTATGTAGTACAGTATCTCCATGTACCTG 780  
OY 781 caaccacacttgcagcagctcccaaggaacaaagaatggttctctcttctatggaac 840  
DB 781 CAACCACTTTCACCCAGCTCCCAAGACCGGGAAGATGTTTCTCTCTGTGTGAATC 840  
OY 841 attgacccatctgaatcccttatatatatacttaggaacaaagaggttaagaagc 900  
DB 841 ATTGCACCCATGCTAATCCCTTATATATACCTTAGACAAAGAGTAAAGAGAGCC 900  
OY 901 tttaaaaggttggttcg-aagagctcttcttaacaaagaataa 942  
DB 901 TTTAAAGGTTGTTGCAAGAGCTCTTATATCAAGAAATTA 943

RESULT 13  
HS193B12/C  
LOCUS

## DEFINITION

Human DNA sequence from PAC 193B12 on chromosome 6p21.3-22.3. Contains histone H2A/d, H2B/d, H2A.1, H1.5, H3.F, H4.K, H3/J genes, histone PH2B.1 and hypothetical protein H4 pseudogenes, histone H2A.1 and H2B.2 duplicate genes, Glycine (Ggc) tRNA gene, olfactory receptor OLI like gene, ESTs STSs and predicted CpG islands.

## ACCESSION

298744.1 GI:3080457

## VERSION

6p21.3-22.3; A4: C6b island; Glycine tRNA: H1.5; H2A: H2A.1; H2A/d; H2B: H2B.2; H2B/d; H3: H3.1; H3.F; H3/J; H4: H4.K; histone; OLI: olfactory receptor; PH2B.1.

## KEYWORDS

human.

## SOURCE

Homo sapiens

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 (bases 1 to 100375)

## AUTHORS

Wild,A.

## TITLE

Direct Submission

## JOURNAL

Submitted (12-DEC-1997) Chromosome 6 Project Group  
(http://www.sanger.ac.uk/HGP/Chr6/) Sanger Centre, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humqueresanger@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Apr 24, 1998 this sequence version replaced gi:2578067.  
IMPORTANT: This sequence is the entire insert of clone 193B12.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variations annotated may not be found in the sequence submission  
corresponding to the overlapping clone as we submit sequences with  
only a small overlap as described above.  
This sequence was generated from part of bacterial clone contigs of  
human chromosome 6, constructed in collaboration by the Sanger  
Centre chromosome 6 mapping group and Armin Volz & Andreas Ziegler,  
David Ruddy, Jeffrey Gruen.  
Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr6/ This sequence has been finished  
according to sequence map criteria as follows. An attempt is made  
to resolve all sequencing problems, such as compressions and  
repeats, but not necessarily within known annotated human repeat  
sequence elements (e.g. Alu). Where the sequence is ambiguous,  
there is an annotation using the 'unsure' feature key.  
The true left end of clone 193B12 is at 1 in this sequence. The

## COMMENT

95T/CY 0TC/10 T0EECY

	/codon_start=1	
	/evidence-not_experimental	
misc_feature	27829..29369	
	/note="putative Cpg island"	
repeat_region	28190..28492	
	/note="AlusX repeat: matches 1..302 of consensus"	
repeat_region	28508..28807	
	/note="AlusQ repeat: matches 2..303 of consensus"	
mRNA	<29070..>29351	
	/gene="H2A.1"	
Query Match	77.1%; Score 726.2; DB 9; Length 100375;	
Best Local Similarity	86.3%; Pred. No. 3.8e-197;	
Matches 814; Conservative	0; Mismatches 128; Indels 1; Gaps	1;
Oy	1 atgaattcggtaataagacacatcacagaagattatctctcgtggttcctcaatcgca	60
Dd	76061 ATGAAATTGGGTAAATAAGAGTGTCCACAGSAGTTTCATCTTGTTAGTTTCACATCA	76002
Oy	61 ccttgctcggagtttccacctctgttgtgtctcttgattcttacaactgtgcaccatctt	120
Dd	76001 CCATGGCTAGAGAAATTCACCACCTTTGTGAATGTTTCGTTTCCTAATATCTTGCAAAATCTTT	75942
Oy	121 ggcacatcgcagcatatttatctagtgcagccgcggacacccaacttcataccccatgat	180
Dd	75941 GGCAATCTGCACAATAAATCTTGTCACATGTCGAATTCAAAATCACCACCCCTATGTAC	75862
Oy	181 tttctcttacaaatcatatcaactcctgtgactcttctgtttacacacatatgtaagttccaca	240
Dd	75881 TTTTTCCTTAGCAATCTGCACATCCTCGACACTTTTGCTATACCAAGAATACATTCACAA	75822
Oy	241 atgcctagtaaatctatcgcagcatcaggaaaataatcagttalcgtgctgtgtagccag	300
Dd	75821 ATGCTGCTAAACATATATGCAMCACACAGAAAGTAATCATGTTATGTCGGCTGTGGCCCCAG	75762
Oy	301 ctcttcataattctcggccttggggcctctctgataatctctctccggccgaatgctcctt	360
Dd	75761 CTTTCATATTTCTCTGGCCTTGGGCTTCACAGAAATGCTTCTCTGGCCGCTCATGTGCTTT	75702
Oy	361 gataagttctgatactatctgtcgcgcctccatctaactcagttatcatgatcacaagaatc	420
Dd	75701 GATAGTGTGTATGATCTATTGTGTGGGCTGTCCATTTACTCAATATATCATGACCAAGAGCTC	75642
Oy	421 tgcctccagttvgcagccgacatcctgygttactggtltagtiaactcaagtgtgtgtct	480
Dd	75641 TGCTTCCAGTGGGACGTGCATCCAGTATGATGGCTTTAGCAATTCAGATTTAACAGTCC	75582
Oy	481 acccgcatctccagctctgcagctcagctgtgaacctatgatatgaatcacttctctgttaa	540
Dd	75581 ACCTGACACCTTAATATGCCACTGTGTGGTCCAAGAAAGATGATGACTTCTTCTGTGAA	75522
Oy	541 gtcccctgcagctcgaagttaactctgtgtgttgagacaacagcaaatgaagctgaactatc	600
Dd	75521 GTCCCTGCTCTGCTCAAGTGTCTGTGTGACACAAACGCAAATGAGGCTAACTATTTC	75462
Oy	601 ctgttcagltgagctcttccatctaaataacccctigacactatccctatatcatatgcitlt	660
Dd	75461 TTCAATCAGTGTGCTATTCCTCTTATATACCCGAGACACTCATCTTATATCATGATGCTTTT	75402
Oy	661 attgctcggagcagatctatgagataaagctctgtcgaagtcgcacaaaagcatltyggaca	720
Dd	75401 ATTGTTCAGAGCGTGTGTGAGAAATTCAGTCTGCTGAAGGTCAACGAAAGGCATTTTGGACA	75342
Oy	721 ttgtgttcccatctaattgtgtgtctctcttctttagtaagacgcgtctcctgtfaccctg	780
Dd	75341 TGTGGCTCCCATCTAATGTGTGTGTCACCTTTTATATGTAACGCTATCTCCATGTAACTTG	75282
Oy	781 caaacacacttcgccagctcacaagaccaaaggaagaatggttctctctcttcataatgaac	840
Dd	75281 CAACCAACGCTTACCCAGGCTCCAAAGACCGGGGAAAGATGGTTTCTCTCTTCTGTGGAAATC	75222
Oy	841 attgcaccaatgcgtgaatccctctatatatacatctttaggaacaaaggaagtlaagaaagcc	900

QY	901	TTTAAAGTGTGGTTCGAAGAGCTCTCTTAATCAAGAATAATAA	75119
DB	75161		
RESULT	14		
LOCUS	AC024428		
DEFINITION	AC024428	166758 bp	DNA
ACCESSION	AC024428		HTG
VERSION	AC024428.3		26-MAY-2000
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryotic; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
TITLE	1 (bases 1 to 166758)		
JOURNAL	Birten, B., Linton, L., Nusbaum, C. and Lander, E.		
REFERENCE	Homo sapiens chromosome 3, clone RP11-59801		
AUTHORS	Unpublished		
	2 (bases 1 to 166758)		
	Birten, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,		

TITLE	Direct Submission
JOURNAL	Submitted (28-FEB-2000), Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On May 23, 2000 this sequence version replaced gi:7212173.

```

-----
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: f6934
Center clone name: 598_O-1
-----
Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 156318 bases at least Q40
Consensus quality: 161224 bases at least Q30
Consensus quality: 163451 bases at least Q20
Insert size: 191000; agarose-fp
Insert size: 165158; sum-of-contigs
Quality coverage: 4.0 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-contigs
-----

```

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 17 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```
1 1025: contig of 1025 bp in length
1026 1125: gap of 100 bp
1126 3025: contig of 1900 bp in length
3026 3125: gap of 100 bp
3126 4346: contig of 1221 bp in length
4347 4446: gap of 100 bp
4447 7420: contig of 2974 bp in length
7421 7520: gap of 100 bp
7521 10669: contig of 3149 bp in length
10670 10769: gap of 100 bp
10770 14258: contig of 3489 bp in length
14259 14358: gap of 100 bp
14359 18958: contig of 4600 bp in length
18959 19058: gap of 100 bp
19059 23968: contig of 4910 bp in length
23969 24068: gap of 100 bp
24069 28546: contig of 4478 bp in length
28547 28646: gap of 100 bp
28647 37679: contig of 9033 bp in length
37680 37779: gap of 100 bp
37780 45607: contig of 7828 bp in length
45608 45707: gap of 100 bp
45708 54439: contig of 8732 bp in length
54440 54539: gap of 100 bp
54540 63730: contig of 9191 bp in length
63731 63830: gap of 100 bp
63831 77842: contig of 14012 bp in length
77843 77942: gap of 100 bp
77943 98659: contig of 20717 bp in length
98660 98759: gap of 100 bp
98760 127843: contig of 29084 bp in length
127844 127943: gap of 100 bp
127944 166758: contig of 38815 bp in length.
```

## FEATURES

```
source
1. 166758
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-59801"
/clone_id="RP11-59801"
/clone_id="RP11-59801"
misc_feature
1. 1025
/note="assembly_fragment"
misc_feature
1126. 3025
/note="assembly_fragment"
misc_feature
3126. 4346
/note="assembly_fragment"
misc_feature
4447. 7420
/note="assembly_fragment"
misc_feature
7521. 10669
/note="assembly_fragment"
misc_feature
10770. 14258
/note="assembly_fragment"
misc_feature
14359. 18958
/note="assembly_fragment"
misc_feature
19059. 23968
/note="assembly_fragment"
misc_feature
24069. 28546
/note="assembly_fragment"
misc_feature
45608. 45707
/note="assembly_fragment"
misc_feature
54440. 54539
/note="assembly_fragment"
misc_feature
63731. 63830
/note="assembly_fragment"
misc_feature
77843. 77942
/note="assembly_fragment"
misc_feature
98660. 98759
/note="assembly_fragment"
misc_feature
127844. 127943
/note="assembly_fragment"
```

```
misc_feature 37780..45607
/note="assembly_fragment"
misc_feature 45708..54439
/note="assembly_fragment"
misc_feature 54540..63730
/note="assembly_fragment"
misc_feature 63831..77842
/note="assembly_fragment"
misc_feature 77943..98659
/note="assembly_fragment"
misc_feature 98760..127843
/note="assembly_fragment"
misc_feature 127944..166758
/note="assembly_fragment"
BASE COUNT 48778 a 33823 c 33960 g 48591 t 1606 others
ORIGIN
Query Match 77.1%; Score 726.2; DB 2; Length 166758;
Best Local Similarity 86.3%; Pred. No. 4e-197;
Matches 814; Conservative 0; Mismatches 128; Indels 1; Gaps 1;
QY 1 atgaattggtaaatgacagatcatatagagagatttctcgtggttctcagatga 60
|||||
Pb 99340 ATGAATTGGTAATAATGAGAGTCCACAGAGATTCATTGTTGTTCCAGATCAA 99399
QY 61 ccttggtcgtggttccacccctcctggtgtcttctgatttcttaacgtgacatctt 120
|||||
Pb 99400 CCATGCGTAGAGATGCCACCCCTTGTGATGTTCTGTTCTTATATCTTGACAAATCTTT 99459
QY 121 ggaactgacatattctagtcagcgcctgacacacacatcatcaccacatgat 180
|||||
Pb 99460 GGAATCTGACATATATCTTGTGTGACATGTGGATTTTCAAACTCACACCCCTATGTAC 99519
QY 181 ttttcttacaatcatcatcctcctgtagcttctgttaacacacatgtacagttccacaa 240
|||||
Pb 99520 TTTTCTTACCAATCTCTCACTCTGACCTTCTCTATACCAAGTACAGTTCCACAA 99579
QY 241 atgcagtaaatattatgacagatcaggaagaatgaattatcggtgtgtgagccag 300
|||||
Pb 99580 ATGCTGATTAACATATGTAACACCCAGGAAGTAAATCACTTATGAGTGTGTGCGCCAG 99639
QY 301 ctttcatattcttgaccttggtggtggtggtggtggtggtggtggtggtggtggt 360
|||||
Pb 99640 CTTTTCATTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 99699
QY 361 gatagtttgtagctatttgcgcctccatcatcagttatcatatgacagagagact 420
|||||
Pb 99700 GATAGTTTGTGATGATTTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 99759
QY 421 tgcctccagttggcagccgcatccctggttactggttttagtaactcagtggtgtct 480
|||||
Pb 99760 TGCCTCCAGTTGGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 99819
QY 481 accctgactccagctgcacactcgtgaccccttgatagatgacacttctcgtgaa 540
|||||
Pb 99820 ACCCTGACTCTTAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 99879
QY 541 gtcctgacagctcaagttatcttctgtgtgagacaagacaagaaatgaggtgacatctc 600
|||||
Pb 99880 GTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 99939
QY 601 ctgtcagtgagctcttccatcatatccctggaacatccatccatcatatcatatctt 660
|||||
Pb 99940 TTCATCAGTGTCTATTCCTCTTATATACCCGTCGACATCTTATATCTTATCTTAT 99999
QY 661 attgctcagagatattagagatagacagctgtagaagtcgacaaagaacatttgagaa 720
|||||
Pb 100000 ATTGCTCAAGCAGCTGTTGAGATTCAGTCTGTGAGAGTCACAGCAAGCAATTTGGACA 100059
QY 721 tgtgttccacataattgtgtgtctctttttagtagaacagcgtctcgtgtgactgt 780
|||||
Pb 100060 TGTGCTCCCATCTAATGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 100119
```

```

QY 781 caaccacttcgcccagctcccaagaccagaagaatgcttcctcctcatgaatc 840
      ||||||| ||||||| ||| ||||||| ||||||| |||||||
Db 100120 CAACACCTTACACCGACGTCACCAAGACCGGAGAAAGATGTTCTCTCTGTGAAATC 100179
      ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 841 attgcacacatcgtgaatcccttatatatacacttaagacaagaggtaaagaagc 900
      ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 100180 ATTGCACCCCTGCTGATCCCTTATATATACACTTAGACAAAGAGCTAAAGCAAGCC 100239
      ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 901 tttaaaagttgttgc-aagagctcttctaagaataa 942
      ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 100240 TTAAAAAGGTGTCACAAAGCTTCTTAATCAGAAATA 100282
      ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 15
AC025941
LOCUS AC025941 176277 bp DNA HTG 26-MAY-2000
DEFINITION Homo sapiens chromosome 6 clone RP11-635011 map 6, WORKING DRAFT
SEQUENCE, 15 unordered pieces.
ACCESSION AC025941
VERSION AC025941.2 GI:7408057
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 176277)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
JOURNAL Homo sapiens chromosome 6, clone RP11-635011
REFERENCE 2 (bases 1 to 176277)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F.,
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
Campotiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collins, R., Cooke, P., Deavello, K., Dewar, K., Diaz, J. S.,
Dodg, S., Dominko, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kamp, L., Karatas, A.,
Klein, R., Lahocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,
Medlitz, M., Meneus, L., Mihova, T., Miranda, C., Mlenka, V., Morrow, J.,
Murphy, T., Naylor, T., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (17-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 4, 2000 this sequence version replaced gi:7259782.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8577
Center clone name: 635-O_11
----- Summary Statistics
Sequencing vector: M13; M77815: 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 168929 bases at least Q40
Consensus quality: 172827 bases at least Q30

```

```

Consensus quality: 174151 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 174877; sum-of-contigs
Quality coverage: 4.5 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 2997: contig of 2997 bp in length
2998 3097: gap of 100 bp
3098 5606: contig of 2509 bp in length
5607 5706: gap of 100 bp
5707 9209: contig of 3503 bp in length
9210 9309: gap of 100 bp
9310 13614: contig of 4305 bp in length
13615 13714: gap of 100 bp
13715 20639: contig of 6925 bp in length
20640 20739: gap of 100 bp
20740 26545: contig of 5806 bp in length
26546 26645: gap of 100 bp
26646 37068: contig of 10423 bp in length
37069 37168: gap of 100 bp
37169 48083: contig of 10915 bp in length
48084 48183: gap of 100 bp
48184 59015: contig of 10832 bp in length
59016 59115: gap of 100 bp
59116 70420: contig of 11305 bp in length
70421 70520: gap of 100 bp
70521 80466: contig of 9946 bp in length
80467 80566: gap of 100 bp
80567 92325: contig of 11759 bp in length
92326 92425: gap of 100 bp
92426 110572: contig of 18147 bp in length
110573 110672: gap of 100 bp
110673 131898: contig of 21226 bp in length
131899 131998: gap of 100 bp
131999 176277: contig of 44279 bp in length.
Location/Qualifiers
1. 176277
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6"
/clone_id="RP11-635011"
/clone_idb="RPC1-11 Human Male BAC"
1. 2997
/note="assembly_fragment"
3098 .5606
/note="assembly_fragment"
clone_end:sp6
vector_side:left"
5707 .9209
/note="assembly_fragment"
9310 .13614
/note="assembly_fragment"
13715 .20639
/note="assembly_fragment"
20740 .26545
/note="assembly_fragment"
clone_end:r7
vector_side:right"
26646 .37068
/note="assembly_fragment"
37169 .48083
/note="assembly_fragment"
48184 .59015
/note="assembly_fragment"

```

```

misc_feature      59116..70420
                   /note="assembly_fragment"
misc_feature      70521..80466
                   /note="assembly_fragment"
misc_feature      80367..92325
                   /note="assembly_fragment"
misc_feature      92426..110572
                   /note="assembly_fragment"
misc_feature      110673..131898
                   /note="assembly_fragment"
misc_feature      131999..176277
                   /note="assembly_fragment"
BASE COUNT      54493 a 33357 c 33254 g 53772 t 1401 others
ORIGIN

```

```

Query Match      77.18; Score 726.2; DB 2; Length 176277;
Best Local Similarity 86.38; Pred. No. 4e-197;
Matches 814; Conservative 0; Mismatches 128; Indels 1; Gaps 1;

QY 1 atgaattggtaaaagacagcatcatatagagagattatcctggtgttccagatcga 60
   |||||
Db 41072 ATGAATTTGGGTAATAAGAGTGCTCCACAGAGGATTCATCTGTAGTTTTCAGATCAA 41131

QY 61 ccttgctgaggttccactccttggtgttcttctgattcttaccactgagcatctt 120
   |||||
Db 41132 CCATGGCTAGAGATTCACCCCTTTGTGATGTTCTGTCTTCTATATCTTGCAATCTTT 41191

QY 121 ggcacatcgacatattatctagtgacgcctgagacacacatcatacccatglat 180
   |||||
Db 41192 GGCATATCGACAAATTAATCTTGTCACATGATGATTCAAACTCCACACCCCTATGTAC 41251

QY 181 ttttttaccacatctactcctctgatacttltgttaccacacatgtacagtcacca 240
   |||||
Db 41252 TTTTTCCTTACCAATCTCTGCTCTGACCTCTGATACCAAGTACCAAGTACCAAGTACCA 41311

QY 241 atgctagtaaaattatagcagcatcaggaagaatlaacagttatcgtgctgtgaccca 300
   |||||
Db 41312 ATGCTGGTAACATATGCAACACGAAAGATTAATAGTATGCTGTGCTGTGCTGTG 41371

QY 301 ctttcatacttctgaccttggggctactgaatacttctctcgtgacctgtcctt 360
   |||||
Db 41372 CTTTTCATTTCTCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 41431

QY 361 gataggtttagatctatttgcgcctctccatctacgtatcatatgacacagaactc 420
   |||||
Db 41432 GATAGGTTTGTAGCTATTGTTGGGCTCTCATTAATCATATGACACAGAGGCTC 41491

QY 421 tggccctcagttggcagcgcatcctgggttactggttttagtaactcagtggtgct 480
   |||||
Db 41492 TGCTTTCAGTTGGCAGCTGCATCTGGAATTAAGGCTTTAGCAATTCAGTATACAGTTC 41551

QY 481 accctgactccacagctgacactctgtgacccctatgtatagatcacttctctgtgaa 540
   |||||
Db 41552 ACCGTGACACTTAAGATGCCACTGTGTGTGCACAAAGAGTGATCATTCTTGTGAA 41611

QY 541 gtccctgacgtcacaattatctgtgtgttagacacacagcaaatgaggctgaactatc 600
   |||||
Db 41612 GTCCCTGCTGTGCTCAAGTTGCTGTGTGTGACACACAGCAAAATGAGGCTGAACATAATTC 41671

QY 601 ctgtcagtgagctcttcatcattacccctgagacatcctcatatcatatgctttt 660
   |||||
Db 41672 TTTCATCATGTGCTATTCTTCTTATATACCGTGACACTCATCTTATATCGATGCTTTT 41731

QY 661 attgctcagagcaglatltagagatacagtcctgtgaagtcgacaaaagcatttgggaca 720
   |||||
Db 41732 ATTGCTCAAGCAGGTGTGAGATCCAGTCTGTGAAGTCACAAAGGCAATTTGGGACA 41791

QY 721 tgtggttcccatctaatgtgtgtctcttlttatagtacagccgtctgtgtactctg 780
   |||||
Db 41792 TGTGCTCCCATCTTAATATGTGTGTGCTCTTTTATGTGTACAGCTATCTCCATGTACTGTG 41851

QY 781 caaccaccttgccacagctccaagcaagaagaagatgttctctctctatgtaatc 840

```

```

|||||
Db 41852 CAACACCTTTCACCCAGCTCCCAAGACCGGGCAAGATGTTTCTCTCTGTGTGATTC 41911

QY 841 attgacccatgctgaatcccttatatacatcttagaacaagagtgtaaggaagc 900
   |||||
Db 41912 ATTGACCCCATGCTGAATCCCTTATATATACCTTAGAACAAAGAGGTAAGGAGACC 41971

QY 901 tttaaaggttggtgctc-aagagctctcttaatacaagaataa 942
   |||||
Db 41972 TTTAAAGGTTGTTGCAAAAGAGTCTTATCAAGAAATAA 42014

```

Search completed: February 26, 2002, 05:04:54  
 Job time: 5022 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 26, 2002, 04:13:22 ; Search time 114.05 Seconds

(Without alignments)  
7081.108 Million cell updates/sec

Title: US-09-755-017-1

Sequence: 1 atgaattggtaaatgacag.....tcttcttaacgaagaataa 942

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N\_Geneseq\_1101:\*

1:	/SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2:	/SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3:	/SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4:	/SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5:	/SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6:	/SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7:	/SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8:	/SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9:	/SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10:	/SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11:	/SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12:	/SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13:	/SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14:	/SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15:	/SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16:	/SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17:	/SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18:	/SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19:	/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20:	/SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21:	/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22:	/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	942	100.0	942	22	AA508541	DNA encoding novel
2	942	100.0	1488	22	AA508542	ORF encoding novel
3	937.4	99.5	939	22	AAH32018	Human olfactory re
4	726.2	77.1	1071	22	AAH31669	Human olfactory re
5	726.2	77.1	1071	22	AAH32250	Human olfactory re
6	641.6	68.1	648	22	AAH32231	Human olfactory re
7	537.2	57.0	1442	21	AACT7475	Human ORF3030
8	534.2	56.7	939	22	AAH32304	Human olfactory re
9	400.4	42.5	936	22	AAH31645	Human olfactory re
10	370	39.3	1002	22	AAH32063	Human olfactory re
11	364.2	38.7	1011	22	AAH32306	Human olfactory re

12	356.6	37.9	933	22	AAH32308	Human olfactory re
13	356	37.8	936	22	AAH32314	Human olfactory re
14	354	37.6	927	22	AAH32067	Human olfactory re
15	352.4	37.4	936	22	AAH32045	Human olfactory re
16	350.2	37.2	1014	22	AAH32316	Human olfactory re
17	350.2	37.2	1440	21	AACT6960	Human ORF2515
18	341.6	36.3	936	22	AAH31632	Human olfactory re
19	341.2	36.2	979	22	AAH32055	Human olfactory re
20	335.8	35.6	909	22	AAH32043	Human olfactory re
21	332.6	35.3	1629	22	AAH32043	Human olfactory re
22	328	34.8	1002	22	AAH32238	Human novel 32164
23	327.2	34.7	948	22	AAH31841	Human olfactory re
24	327.2	34.7	948	22	AAH32375	Human olfactory re
25	322.4	34.2	948	22	AAH32188	Human olfactory re
26	318	33.8	954	22	AAH31588	Human olfactory re
27	318	33.8	957	22	AAH32072	Human olfactory re
28	318	33.8	957	22	AAH32088	Human olfactory re
29	315.6	33.5	1450	21	AACT7005	Human ORF2560
30	314.4	33.4	1087	22	AAH32312	Human olfactory re
31	313.6	33.3	817	22	AAH32390	Human olfactory re
32	311	33.0	933	22	AAH31783	Human olfactory re
33	308.6	32.8	960	22	AAH32023	Human olfactory re
34	306.2	32.5	971	22	AAH32318	Human olfactory re
35	300.8	31.9	1014	22	AAH31672	Human olfactory re
36	294.6	31.3	948	22	AAH32437	Human olfactory re
37	293.6	31.2	923	22	AAH31878	Human olfactory re
38	291.2	30.9	997	22	AAH32019	Human olfactory re
39	289.8	30.8	810	22	AAH32174	Human olfactory re
40	283	30.0	921	22	AAH35759	Human partial olfa
41	282.2	30.0	951	22	AAH31879	Human olfactory re
42	280.6	29.8	930	22	AAH35755	Human olfactory re
43	280.6	29.8	951	22	AAH31584	Human olfactory re
44	278.2	29.5	1030	22	AAH31633	Human olfactory re
45	277.2	29.4	1068	22	AAH31630	Human olfactory re

#### ALIGNMENTS

RESULT 1  
AA508541  
ID AA508541 standard; cDNA; 942 BP.  
XX  
AC AA508541;  
XX  
XX  
XX 26-SEP-2001 (first entry)  
XX  
XX  
DE DNA encoding novel human G-protein coupled receptor (NGPCR).  
XX  
XX Human: novel G-protein coupled receptor; NGPCR; drug discovery;  
KW diagnostic; ss.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX CDS 1..942  
XX FT /\*tag= a  
XX FT /product= "Novel G-protein coupled receptor (NGPCR)"  
XX  
XX MO200151634-A1.  
XX PD 19-JUL-2001.  
XX  
XX 05-JAN-2001; 2001WO-US00589.  
XX PF 12-JAN-2000; 2000US-0175764.  
XX PR (LEXI-) LEXICON GENETICS INC.  
XX PA  
XX Walke DW, Wilganowski NL, Turner CA, Friedrich G, Abuin A;  
XX Zambrowicz B, Sands AT;  
XX WPI: 2001-442145/47.

DR P-PSDB: AAU04689.

XX New human G protein-coupled receptor and polynucleotides encoding the  
PT receptor, useful in identifying, selecting or validating new molecular  
PT targets for drug discovery and in diagnostic or prognostic assays

PS Claim 1: Page 61: 65pp: English.

CC The sequence represents the coding sequence of novel human G-protein  
CC coupled receptor (NGPCR). The NGPCR nucleic acid and protein are useful  
CC in identifying, selecting or validating novel molecular targets for drug  
CC discovery, as well as in diagnostic or prognostic assays. These are also  
CC useful in microarrays or other assay formats, for screening collections  
CC of genetic material from patients who have a particular medical condition  
CC or for identifying mutations associated with a particular disease.

XX Sequence 942 BP: 222 A; 228 C; 187 G; 305 T; 0 other;

Query Match 100.0%; Score 942; DB 22; Length 942;

Best Local Similarity 100.0%; Pred. No. 1,4e-279; Mismatches 0; Gaps 0;

Matches 942; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 atgaattgggaatgaagcatcatagagagattatctgtgtgttcttcagatcga 60
DB 1 atgaattgggaatgaagcatcatagagagattatctgtgtgttcttcagatcga 60
GY 61 ccttggtggaggttcctctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 120
DB 61 ccttggtggaggttcctctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 120
OY 121 ggaactgcagcatatctatctagtcagcgcctgcagacccaactcataccatgta 180
DB 121 ggaactgcagcatatctatctagtcagcgcctgcagacccaactcataccatgta 180
OY 121 ggaactgcagcatatctatctagtcagcgcctgcagacccaactcataccatgta 180
DB 121 ggaactgcagcatatctatctagtcagcgcctgcagacccaactcataccatgta 180
OY 181 tttttcttaccatctatctactcctgtgattctgttacaacacatgtaagttccaca 240
DB 181 tttttcttaccatctatctactcctgtgattctgttacaacacatgtaagttccaca 240
OY 241 atgcagtaaatatttgagacatcagagaaagtaacagttatcggtctgtgtacccag 300
DB 241 atgcagtaaatatttgagacatcagagaaagtaacagttatcggtctgtgtacccag 300
OY 301 ctcttcataattctgtgcttggggctactcgtatatactctcctgcgcgtacgtcctt 360
DB 301 ctcttcataattctgtgcttggggctactcgtatatactctcctgcgcgtacgtcctt 360
OY 361 gataagtttgaagcattctgtgcgcctcctcatcactcaagttatcatgacagagact 420
DB 361 gataagtttgaagcattctgtgcgcctcctcatcactcaagttatcatgacagagact 420
OY 421 tggccctcaggttggcagcgcctcctggttactgttttagtaactcgtgtgtgtct 480
DB 421 tggccctcaggttggcagcgcctcctggttactgttttagtaactcgtgtgtgtct 480
OY 481 accctgactctcagcagctgcacactcgtgacccctatgtagatagacactttctcgtaa 540
DB 481 accctgactctcagcagctgcacactcgtgacccctatgtagatagacactttctcgtaa 540
OY 541 gtccctcagcagctcgaatctatctgtgtgtgagacaagcaaatgaggtgtaactatc 600
DB 541 gtccctcagcagctcgaatctatctgtgtgtgagacaagcaaatgaggtgtaactatc 600
OY 601 ctgtgcagtgagctcttcatactaataccctgcagacatccatccatatacatgtcttt 660
DB 601 ctgtgcagtgagctcttcatactaataccctgcagacatccatccatatacatgtcttt 660
OY 661 attgctcagcagatctgagatacagctctgtaaggtgcacaaaagcatctggagaca 720
DB 661 attgctcagcagatctgagatacagctctgtaaggtgcacaaaagcatctggagaca 720
OY 721 tctgttccatctaatgtgtgtctcttcttcttcttcttcttcttcttcttcttctt 780
DB 721 tctgttccatctaatgtgtgtctcttcttcttcttcttcttcttcttcttcttctt 780

```

```

DB 721 tctgttccatctaatgtgtgtctcttcttcttcttcttcttcttcttcttcttctt 780
OY 781 caaccacttcggccagctcgaagacccaaggaagatgttctcttcttcttcttcttctt 840
DB 781 caaccacttcggccagctcgaagacccaaggaagatgttctcttcttcttcttcttctt 840
OY 841 attgacccatgcgtatcccttatactaataggaaggaaggaaggaaggaaggaaggaag 900
DB 841 attgacccatgcgtatcccttatactaataggaaggaaggaaggaaggaaggaaggaag 900
OY 901 tttaaaagttgtgtcagaagctcttcttcttcttcttcttcttcttcttcttcttctt 942
DB 901 tttaaaagttgtgtcagaagctcttcttcttcttcttcttcttcttcttcttcttctt 942

```

RESULT 2

AS08542 standard: CDNA: 1488 BP.

AS08542:

26-SEP-2001 (first entry)

ORF encoding novel human G-protein coupled receptor (NGPCR).

Human: novel G-protein coupled receptor: NGPCR; drug discovery;

diagnostic; ss.

Homo sapiens.

Key Location/Qualifiers

FT CDS 198..1139

FT /tag= a /product= "Novel G-protein coupled receptor (NGPCR)"

MO200151634-A1.

19-JUL-2001.

05-JAN-2001; 2001WO-US00589.

12-JAN-2000; 2000US-0175764.

(LEXI-) LEXICON GENETICS INC.

Walke DM, Wilganowski NL, Turner CA, Friedrich G, Abuin A;

Zambrowicz B, Sands AT;

WPI: 2001-442145/47.

P-PSDB: AAU04689.

New human G protein-coupled receptor and polynucleotides encoding the  
PT receptor, useful in identifying, selecting or validating new molecular  
PT targets for drug discovery and in diagnostic or prognostic assays

PS Disclosure: Page 62-63: 65pp: English.

CC The sequence represents the open reading frame (ORF) encoding novel human  
CC G-protein coupled receptor (NGPCR). The NGPCR nucleic acid and protein  
CC are useful in identifying, selecting or validating novel molecular  
CC targets for drug discovery, as well as in diagnostic or prognostic  
CC assays. These are also useful in microarrays or other assay formats, for  
CC screening collections of genetic material from patients who have a  
CC particular medical condition or for identifying mutations associated with  
CC a particular disease.

XX Sequence 1488 BP: 421 A; 333 C; 254 G; 477 T; 3 other;

Query Match 100.0%; Score 942; DB 22; Length 1488;

Best Local Similarity 100.0%; Pred. No. 1,7e-279;

Matches 942; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DE	Human olfactory receptor polynucleotide, SEQ ID NO: 591.
XX	
KW	human; olfactory receptor; OR; primary scent determination;
KW	secondary scent determination; polypeptide library; odour receptor;
KW	scent profile; scent fingerprint; scent representation; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200127158-A2.
PD	19-APR-2001.
PP	06-OCT-2000; 2000WO-US27582.
PR	08-OCT-1999; .99US-0158615.
PR	24-FEB-2000; 2000US-0184809.
PA	(DIGI-) DIGISCENTS.
PA	(YEDA ) YEDA RES & DEV CO LTD.
PI	Bellenson J, Smith D, Lancel D, Glusman G, Fuchs T, Yanai I;
DR	WPI; 2001-290713/30.
PT	New polynucleotides which encode polypeptides involved in olfactory
PT	sensation for identifying olfactory agonists and antagonists -
FS	Claim 8; Page 417; 1857pp; English.
XX	
CC	The present sequence is one of a number of isolated polynucleotides
CC	which encode polypeptides involved in olfactory sensation. The
CC	polynucleotides can be used in screening for olfactory agonists and
CC	antagonists. The methods allow for the determination of primary
CC	scents and the identification of the odour receptors used to detect
CC	these primary scents. The methods also enable determination of
CC	secondary scents and the identification of combinations of odour
CC	receptors that are involved in detecting such secondary scents.
CC	This enables the construction of a scent representation (also called
CC	a scent fingerprint or scent profile), which may be used to re-create
CC	and edit scents. Libraries of olfactory receptors are useful for
CC	determining the interaction pattern of a composition with the receptors,
CC	and can be used for determining differences in the olfactory faculties
CC	of different individuals.
SQ	
	Sequence 939 BP; 220 A; 227 C; 187 G; 305 T; 0 other:
	Query Match            99.5%; Score 937.4; DB 22; Length 939;
	Best Local Similarity   99.9%; Pred. No. 3.5e-278;
	Matches 938; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY	1 atgaattgggtaaaagaagacatcatcaaggagtattctgcgttggttccagatga 60
DB	1 atgaattgggtaaaagaagacatcatcaaggagtattctgcgttggttccagatga 60
OY	61 ccttgccgcggagttcccacctccttcgttgtctcttgatcttctaaccgtgacaactt 120
DB	61 ccttgccgcggagttcccacctccttcgttgtctcttgatcttctaaccgtgacaactt 120
OY	121 ggcacatcgacatattctagtgtagcgcctgagaccaccaactcataccccatgat 180
DB	121 ggcacatcgacatattctagtgtagcgcctgagaccaccaactcataccccatgat 180
OY	181 tttttcttaccaatactatcaactccttgatatttctaacacacatgacagtcccaaa 240
DB	181 tttttcttaccaatactatcaactccttgatatttctaacacacatgacagtcccaaa 240
OY	241 atgcagtaaatatttagcagcatcgaggaagaataacccagttatcgtgagctgttagccag 300
DB	241 atgcagtaaatatttagcagcatcgaggaagaataacccagttatcgtgagctgttagccag 300
OY	301 ctctcacatatctctgaccttgaggagctactgaatactctctcctgagccgatgctctt 360
DB	301 ctctcacatatctctgaccttgaggagctactgaatactctctcctgagccgatgctctt 360

DB 301 ctttccatattctcgtccttgagggtactgaataatcttccctggcgcgtatctctt 360  
QY 361 gataagttgtatgataattgttcggcctccattactacgtatcatcgtaccagaagatc 420  
DB 361 gataagttgtatgataattgttcggcctccattactacgtatcatcgtaccagaagatc 420  
QY 421 tgcctccagttggcagcgcgacatcctgttactgtgtttagtaactcaagtgtgtgtct 480  
DB 421 tgcctccagttggcagcgcgacatcctgttactgtgtttagtaactcaagtgtgtgtct 480  
QY 481 acccgtactcctcagctcgtcagctcgttggccctatgtgatagatcattctctgttga 540  
DB 481 acccgtactcctcagctcgtcagctcgttggccctatgtgatagatcattctctgttga 540  
QY 541 gtcctcgcactgtcgaagtatctgtgttgagacaacagcgaatgagctgaactatc 600  
DB 541 gtcctcgcactgtcgaagtatctgtgttgagacaacagcgaatgagctgaactatc 600  
QY 601 ctgtcagtgagctcttccatcaataacccctgacacatccatccatcatatgctttt 660  
DB 601 ctgtcagtgagctcttccatcaataacccctgacacatccatccatcatatgctttt 660  
QY 661 attgtccgagcaggtttgagatagatcgtctgtgaagtcgacgaacaaagcatttgagaa 720  
DB 661 attgtccgagcaggtttgagatagatcgtctgtgaagtcgacgaacaaagcatttgagaa 720  
QY 721 tctgtgtcccatcaattgtgtgtctctttttatagtagacgctcctgtgtacgt 780  
DB 721 tctgtgtcccatcaattgtgtgtctctttttatagtagacgctcctgtgtacgt 780  
QY 781 caacacactcgcagcagctcgaagcaagcaagaagatgttctctctctatggaac 840  
DB 781 caacacactcgcagcagctcgaagcaagcaagaagatgttctctctctatggaac 840  
QY 841 attgacacactcgtgaatcccttatatatatacacttagaacaagaagtaagaagac 900  
DB 841 attgacacactcgtgaatcccttatatatatacacttagaacaagaagtaagaagac 900  
QY 901 tttaaaaggttggtgcaagagcttcttcaatcaagaa 939  
DB 901 tttaaaaggttggtgcaagagcttcttcaatcaagaa 939

RESULT 4  
AAH31669  
ID AAH31669 standard; DNA: 1071 BP.  
XX  
AC AAH31669;  
XX  
DT 30-JUL-2001 (first entry)  
XX  
DE Human olfactory receptor polynucleotide, SEQ ID NO: 242.  
XX  
KW Human; olfactory receptor; OR; primary scent determination;  
KW secondary scent determination; polypeptide library; odour receptor;  
KW scent profile; scent fingerprint; scent representation; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200127158-A2.  
XX  
PD 19-APR-2001.  
XX  
PF 06-OCT-2000; 2000WO-US27582.  
XX  
PR 08-OCT-1999; 99US-0158615.  
XX  
PR 24-FEB-2000; 2000US-0184809.  
XX  
PA (DIGI-) DIGISCENTS.  
XX  
PA (YEDA) YEDA RES & DEV CO LTD.  
XX  
PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;  
XX

DR WPI: 2001-290713/30.  
XX  
PT New polynucleotides which encode polypeptides involved in olfactory  
PT sensation for identifying olfactory agonists and antagonists -  
XX  
PS Claim 8; Page 273-274; 1857Pp; English.  
XX  
XX  
CC The present sequence is one of a number of isolated polynucleotides  
CC which encode polypeptides involved in olfactory sensation. The  
CC polynucleotides can be used in screening for olfactory agonists and  
CC antagonists. The methods allow for the determination of primary  
CC scents and the identification of the odour receptors used to detect  
CC these primary scents. The methods also enable determination of  
CC secondary scents and the identification of combinations of odour  
CC receptors that are involved in detecting such secondary scents.  
CC This enables the construction of a scent representation (also called  
CC a scent fingerprint or scent profile), which may be used to re-create  
CC and edit scents. Libraries of olfactory receptors are useful for  
CC determining the interaction pattern of a composition with the receptors,  
CC and can be used for determining differences in the olfactory faculties  
CC of different individuals.  
XX

SO Sequence 1071 BP; 277 A; 248 C; 201 G; 345 T; 0 other;

Query Match 77.1%; Score 726.2; DB 22; Length 1071;  
Best Local Similarity 86.3%; Pred. No. 3.4e-213;

Matches 814; Conservative 0; Mismatches 128; Indels 1; Gaps 1;

QY 1 atgaattggglaaagacacacatcatcacaggaattatctcgtgtgttccatcga 60  
DB 1 atgaattggglaaagacacacatcatcacaggaattatctcgtgtgttccatcga 60  
QY 61 ccttggtggaattccactccttctgtgtctcttgattcttaccatgtgacatctt 120  
DB 61 ccttggtggaattccactccttctgtgtctcttgattcttaccatgtgacatctt 120  
QY 121 ggcacatcacacatattcttagtgcagcctgcgacacacacatccatccatgat 180  
DB 121 ggcacatcacacatattcttagtgcagcctgcgacacacacatccatccatgat 180  
QY 181 ttttcttaccacatcactcctcgtgactcttctgttaccacacatgtaagtcacaa 240  
DB 181 ttttcttaccacatcactcctcgtgactcttctgttaccacacatgtaagtcacaa 240  
QY 241 atgcctagtaattatgacgacatcaggaagaatcagttacgtgtgtgtgacgag 300  
DB 241 atgcctagtaattatgacgacatcaggaagaatcagttacgtgtgtgtgacgag 300  
QY 301 ctttccatattctcgtccttgagggtactagatatcttccctggcgcgtacatgctt 360  
DB 301 ctttccatattctcgtccttgagggtactagatatcttccctggcgcgtacatgctt 360  
QY 361 gataagttgtatgataattgttcggcctccattactacgtatcatcgtaccagaagatc 420  
DB 361 gataagttgtatgataattgttcggcctccattactacgtatcatcgtaccagaagatc 420  
QY 421 tgcctccagttggcagcgcgacatcctgttactgtgtttagtaactcaagtgtgtgtct 480  
DB 421 tgcctccagttggcagcgcgacatcctgttactgtgtttagtaactcaagtgtgtgtct 480  
QY 481 acccgtactcctcagctcgtcagctcgttggccctatgtgatagatcattctctgttga 540  
DB 481 acccgtactcctcagctcgtcagctcgttggccctatgtgatagatcattctctgttga 540  
QY 541 gtcctcgcactgtcgaagtatctgtgttgagacaacagcgaatgagctgaactatc 600  
DB 541 gtcctcgcactgtcgaagtatctgtgttgagacaacagcgaatgagctgaactatc 600  
QY 601 ctgtcagtgagctcttccatcaataacccctgacacatccatccatcatatgctttt 660  
DB 601 ctgtcagtgagctcttccatcaataacccctgacacatccatccatcatatgctttt 660

CC of different individuals.

Db 901 tttaaggttggtgcaagagtccttctaatacaagaataa 943

## RESULT 6

ID	AAH32231	standard: DNA; 648 BP.
XX	AAH32231;	
XX	30-JUL-2001 (first entry)	
XX	Human olfactory receptor polynucleotide, SEQ ID NO: 804.	
XX	Human; olfactory receptor; OR; primary scent determination;	
XX	secondary scent determination; polypeptide library; odour receptor;	
XX	scent profile; scent fingerprint; scent representation; ds.	
XX	Homo sapiens.	
XX	MO200127158-A2.	
XX	19-APR-2001.	
XX	06-OCT-2000; 2000WO-US27582.	
XX	08-OCT-1999; 99US-0158615.	
XX	24-FEB-2000; 2000US-0184809.	
XX	(DIGI-) DIGISCENTS.	
XX	(YEDA ) YEDA RES & DEV CO LTD.	
XX	Bellenson J, Smith D, Lancet D, Giusman G, Fuchs T, Yanai I;	
XX	WPI: 2001-290713/30.	
XX	New polynucleotides which encode polypeptides involved in olfactory	
XX	sensation for identifying olfactory agonists and antagonists -	
XX	Claim 8; Page 502; 1857pp; English.	
XX	The present sequence is one of a number of isolated polynucleotides	
XX	which encode polypeptides involved in olfactory sensation. The	
XX	polynucleotides can be used in screening for olfactory agonists and	
XX	antagonists. The methods allow for the determination of primary	
XX	scents and the identification of the odour receptors used to detect	
XX	these primary scents. The methods also enable determination of	
XX	secondary scents and the identification of combinations of odour	
XX	receptors that are involved in detecting such secondary scents.	
XX	This enables the construction of a scent representation (also called	
XX	a scent fingerprint or scent profile), which may be used to re-create	
XX	and edit scents. Libraries of olfactory receptors are useful for	
XX	determining the interaction pattern of a composition with the receptors,	
XX	and can be used for determining differences in the olfactory faculties	
XX	of different individuals.	
XX	Sequence 648 BP; 144 A; 166 C; 132 G; 206 T; 0 other;	
XX	Query Match 68.1%; Score 641.6; DB 22; Length 648;	
XX	Best Local Similarity 99.4%; Pred. NO. 2.8e-187;	
XX	Matches 644; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	
XX	202 ccccgagatcttgttaaccacagtcacgcccacaaatgtcagtaattatgcagc 261	
XX	1 ccccgagatcttgttaaccacagtcacgcccacaaatgtcagtaattatgcagc 60	
XX	262 atcaggaagaatcagttatcgtgtgctgtgagccaaacttccatattctggcctg 321	
XX	61 atcaggaagaatcagttatcgtgtgctgtgagccaaacttccatattctggcctg 120	
XX	322 ggggactactgaatctcttcctggccgctcagtcctcttgatagaagttgtgactattgt 381	
XX	121 ggggactactgaatctcttcctggccgctcagtcctcttgatagaagttgtgactattgt 180	
XX	382 cggccttcctatcattcagttatcagtcacccagaaactctgctccagttggcagccga 441	
XX	181 cggccttcctatcattcagttatcagtcacccagaaactctgctccagttggcagccga 240	

QY	442	ttcttggttacttggtttatagaaactcagtggtgtgtcttaccggtacccgagatctccagcttgcga	501
Db	241	ttccaaaggttacctgggtttttaagtaacctcaatctggtgtgtcttaccggtacccgagatctccagcttgcga	300
QY	502	ctctcttgacccctatgtgtatatagatcaacttctctctgttgaaagctcccttgacatctgctaaagtta	561
Db	301	ctctctgtgacccctatgtgtatagatcaacttctctctgttgaaagctcccttgacatctgctaaagtta	360
QY	562	tctgtgtgttgagacaacaagcaaatgtagagcttgaaactatctctgtctagttgagctcttccat	621
Db	361	tctgtgtgttgagacaacaagcaaatgtagagcttgaaactatctctgtctagttgagctcttccat	420
QY	622	cttaatacccccgaacatcatctcttatatcatatgtcttttatctgtgctcgagagaattatgag	681
Db	421	cttaatacccccgaacatcatctcttatatcatatgtcttttatctgtgctcgagagaattatgag	480
QY	682	atacagctctgtctgaaagctcgacaaaaagcattcttgagacatgtgtctccatcataatgtgtg	741
Db	481	atacagctctgtctgaaagctcgacaaaaagcattcttgagacatgtgtgttccatcataatgtgtg	540
QY	742	gtgtctctcttttllatagtagacgcggtctctgtgtgtaacctgcaaccaacacttcgcccagctcc	801
Db	541	gtgtctctcttttllatagtagacgcggtctctgtgtgtaacctgcaaccaacacttcgcccagctcc	600
QY	802	aaggaaccaaggaagaagtgttctctctcttcatatgtaatacatgtgaccc	849
Db	601	aaggaaccaaggaagaagtgttctctctcttcatatgtaatacatgtgaccc	648

RESULT	7
AAC77475	
ID	AAC77475 standard; cDNA: 1442 BP.
XX	
AC	AAC77475:
XX	
DT	08-FEB-2001 (first entry)
DE	
XX	Human OREF ORF3030 polynucleotide sequence SEQ ID NO:6059.
KW	Human; open reading frame; OREF; detection; cytosolic; hepatotropic;
KW	vulnerable; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
KW	anticonvulsant; osteoprotic; antiarthritis; immunosuppressant; cardiac;
KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW	hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW	antiviral; antibacterial; antifungal; antineumatic; antithyroid;
KW	antihaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW	cholesterol ester storage; systemic lupus erythematosus; infection;
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW	bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW	thrombosis; contraceptive; ss.
OS	Homo sapiens.
XX	
PN	WO200058473-A2.
XX	
PD	05-OCT-2000.
XX	
PF	31-MAR-2000; 2000WO-US08621.
XX	
PR	31-MAR-1999; 99US-0127607.
PR	02-APR-1999; 99US-0127636.
PR	05-APR-1999; 99US-0127728.
PR	30-MAR-2000; 2000US-0540763.
XX	
PA	(CURA-) CURAGEN CORP.
XX	
PI	Shinkets RA, Ibach M;
XX	
WI	WI: 2000-602362/57.
DR	

P-PSDB: AAB43266.

DR Novel nucleic acids and peptides derived from open reading frame X.  
 XX Useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 PS Claim 5; Page 5245-5246; 5507pp; English.

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; valnary;  
 CC antiproliferative; antiparasitoid; neurotropic; neuroprotective;  
 CC osteoplastic; anticonvulsant; antihypertensive; immunosuppressive;  
 CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antineoplastic; antibacterial; antiviral; antifungal; antihypertensive;  
 CC antihypertensive; and antineoplastic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, inflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

SQ Sequence 1442.BP: 378 A; 319 C; 268 G; 477 T; 0 other;

Query Match 57.0%; Score 537.2; DB 21; Length 1442;  
 Best Local Similarity 73.1%; Pred. NO. 5.4e-155;  
 Matches 689; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

QY 1 atgattggtgaatgagacatcatagagattatgctggtggttcagatga 60  
 DB 301 atgattggtgaatgagacatcatagagattatgctggtggttcagatga 360  
 QY 61 cttggtggtggttcctgctgctgctgctgctgctgctgctgctgctgct 120  
 DB 361 gcttggtggtggttcctgctgctgctgctgctgctgctgctgctgctgct 420  
 QY 121 ggcacatgcacatattatgctgctgctgctgctgctgctgctgctgctgct 180  
 DB 421 ggcacatgcacatattatgctgctgctgctgctgctgctgctgctgctgct 480  
 QY 181 ttttttaccacatcctgctgctgctgctgctgctgctgctgctgctgctgct 240  
 DB 481 ttttttaccacatcctgctgctgctgctgctgctgctgctgctgctgctgct 540  
 QY 241 atgctggtgaatgagacatcagagaaatgagacatcagagaaatgagacatc 300  
 DB 541 atgctggtgaatgagacatcagagaaatgagacatcagagaaatgagacatc 600  
 QY 301 ctttaccatcctgctgctgctgctgctgctgctgctgctgctgctgctgct 360  
 DB 601 ctttaccatcctgctgctgctgctgctgctgctgctgctgctgctgctgct 660  
 QY 361 gatggtggtgagcattgctgctgctgctgctgctgctgctgctgctgctgct 420  
 DB 721 gatggtggtgagcattgctgctgctgctgctgctgctgctgctgctgctgct 780  
 QY 481 accctggtggtgagcattgctgctgctgctgctgctgctgctgctgctgctgct 540  
 DB 781 accctggtggtgagcattgctgctgctgctgctgctgctgctgctgctgctgct 840

QY 541 gtccctgagctgctgagacatcattgctgctgctgctgctgctgctgctgctgct 600  
 DB 841 gtccctgagctgctgagacatcattgctgctgctgctgctgctgctgctgctgct 900  
 QY 601 cttgctgagctgctgagacatcattgctgctgctgctgctgctgctgctgctgct 660  
 DB 901 cttgctgagctgctgagacatcattgctgctgctgctgctgctgctgctgctgct 960  
 QY 661 atgctgagcagatgagacatcattgctgctgctgctgctgctgctgctgctgct 720  
 DB 961 atgctgagcagatgagacatcattgctgctgctgctgctgctgctgctgctgct 1020  
 QY 721 tttgctgagcagatgagacatcattgctgctgctgctgctgctgctgctgctgct 780  
 DB 1021 tttgctgagcagatgagacatcattgctgctgctgctgctgctgctgctgctgct 1080  
 QY 781 caaccacatcctgagcagacatcattgctgctgctgctgctgctgctgctgctgct 840  
 DB 1081 caaccacatcctgagcagacatcattgctgctgctgctgctgctgctgctgctgct 1140  
 QY 841 atgctgagcagatgagacatcattgctgctgctgctgctgctgctgctgctgct 900  
 DB 1141 atgctgagcagatgagacatcattgctgctgctgctgctgctgctgctgctgct 1200  
 QY 901 tttgagcagatgagacatcattgctgctgctgctgctgctgctgctgctgct 942  
 DB 1201 tttgagcagatgagacatcattgctgctgctgctgctgctgctgctgctgct 1242

# RESULT 8

AAH32304  
 ID AAH32304 standard; DNA: 939 BP.

AAH32304;  
 30-JUL-2001 (first entry)

Human olfactory receptor polynucleotide, SEQ ID NO: 877.

Human; olfactory receptor; OR; primary scent determination;  
 secondary scent determination; polypeptide library; odour receptor;  
 scent profile; scent fingerprint; scent representation; ds.

Homo sapiens.

WO200127158-A2.

19-APR-2001.

06-OCT-2000; 2000WO-US27582.

08-OCT-1999; 99US-0158615.

24-FEB-2000; 2000US-0184809.

(DIGI-) DIGISENSES.  
 (YEDA ) YEDA RES & DEV CO LTD.

Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

WPI; 2001-290713/30.

New polynucleotides which encode polypeptides involved in olfactory

sensation for identifying olfactory agonists and antagonists -

Claim 8; Page 530; 1857pp; English.

The present sequence is one of a number of isolated polynucleotides  
 which encode polypeptides involved in olfactory sensation. The  
 polynucleotides can be used in screening for olfactory agonists and  
 antagonists. The methods allow for the determination of primary  
 scents and the identification of the odour receptors used to detect  
 these primary scents. The methods also enable determination of  
 secondary scents and the identification of combinations of odour

CC receptors that are involved in detecting such secondary scents.  
CC  
CC This enables the construction of a scent representation (also called  
CC a scent fingerprint or scent profile), which may be used to re-create  
CC and edit scents. Libraries of olfactory receptors are useful for  
CC determining the interaction pattern of a composition with the receptors  
CC and can be used for determining differences in the olfactory faculties  
CC of different individuals.

SQ Sequence 939 BP; 232 A; 228 C; 185 G; 294 T; 0 other;

Query Match	56.7%	Score 534.2	DB 22:	Length 939;
Best Local Similarity	73.1%	Pred. No. 3.6e-154;		
Matches 686;	Conservative 0;	Mismatches 253;	Indels 0;	Gaps 0;

QY	1	atgaaatcgtggaataatgaaagcatatccaaagagcttcttcgctgggtttcttcagatcga	60
Dp	1	atgaaatcgtggaataatgaaagagctccccaagaagatttatactactgtgcttcacgaagi	60
QY	61	cttggctcgtgaggttccactccctctggtgtctctcttgattcttcaactcgtgacacattc	120
Dp	61	gcttggctcacaacaaagcccccttttctgtgtccctgtttaatatatatacaatcacatattt	120
QY	121	ggaacatcgacaacattatctctggttcaagccctgtgaacccaactctacccccatgat	180
Dp	121	ggcaaatgtgtccatcatgatgtgtgtcatctctgtatcccaactctatactcccatgat	180
QY	181	tttttcttacaatctatcaactccctggtatcttgttaacacacatgtaagttccacaa	240
Dp	181	tctcttcccaataatctctcatctttagatctctgtcatatccacaatacagaagtcctcat	240
QY	241	atgctcagaataattatbcagatcataggaaagttaacagtgtaacgttggtctgtgaagccag	300
Dp	241	atgtgtgttaaatatgttgtgcaacaaaagacatcagctatgctgtgctgtgtgtgtccac	300
QY	301	cttttcabattctcgccctctgtgggtcactgtaatatctctcctgtgcgtcatgtctctt	360
Dp	301	ctctacatcttcctcgccctctgagctgtgcaagaagtgtcctctctgtcgtgttatgtcctt	360
QY	361	gataggttttagtattttgttcggccctctccatctacccagttatcatcagccagaagctc	420
Dp	361	gacacgatatgtgtgtgttgcagaagccctccactatgtatgatacatcagaattatgtctc	420
QY	421	tgacctcagttgtgcagccgcacatccctgtggttaactgtgttttagtcaactcagttgtgtct	480
Dp	421	tgccaaagatgtgagccctctcatctggtcatgtgttctgcgaacctcagtgctgcagctc	480
QY	481	acccttgactcccgagcttcgcaactcttgaaaccccatatgtatagatgaactcttctctgaa	540
Dp	481	tcccttgactcttaacatgcgcgcgtgtgtgtacacagaagaatgtgcacacttctctgtgag	540
QY	541	gtccctgcacatgcataagttactctgtgtttgtagacaacagaatgagctgtgacatactc	600
Dp	541	gtgccttgcacatctccaagtgtgtaactgtgcagacaagaacctatgttggtcgtgagctctc	600
QY	601	ctgttcaggttgagccttcacatctataacccctgagcactgaactctctataatattgtctt	660
Dp	601	tctctttagtgaactatctctctataatccagctgtgacatgtgaccccatctccataagctctc	660
QY	661	atgtctcgagagatattgaaagatacagtcctgtgaaggtctgcacaaaagacatttgggaca	720
Dp	661	atagtcacaagatatttaaaaatcagtgctcagcagaagagcgcacaaaagacatttgggaca	720
QY	721	tggtgtcccatctcaatttgggtgtctctcttttatatagataagcgcgtcctgtgtactgt	780
Dp	721	tgtgtgtccccaatgatttgggtgtgtctctctctttaaataagaaacattatattgtatctt	780
QY	781	caaacacccttgcgcagcctcacaagacacaagaagaatgtctctctctctctatgtgactc	840
Dp	781	caaacacccttcatcaccactcagaagatgtgggaagaagtgttctccctctctatgtgactc	840
QY	841	atgcacccatgtctgaatccctctatatatacacttagagaacagaaggttaaaaggaagc	900
Dp	841	atgcacccatgtctgaatccctctatatatacacttagagaacagaaggttaaaaggaagc	900

Db 841 atcaatccacgtttggaactccctatctctacaagcctttgaataataagataagagagagcc 900

Dy 901 tttaaagttggttgcacaagagcttcttcaatcacagaaa 939  
|| || ||| || | |||| |||| ||| |||||

Db 901 ttcaagagctgtagtccaagatctctttctgtctagaaaa 939

RESULT	9
AAH31645	
ID	AAH31645 standard; DNA; 936 BP

DT 30-JUL-2001 (first entry)

Human olfactory receptor polynucleotide, SEQ ID NO: 218.

KW Human; olfactory receptor; OR; primary scent determination;  
 KW secondary scent determination; polypeptide library; odour receptor;  
 KW scent profile; scent fingerprint; scent representation; ds.

OS Homo sapiens.

PN WO200127158-A2.

19-APR-2001.

PE 06-OCT-2000; 2000WO-US27582

PR	08-OCT-1999;	99US-0158615
PR	31-FEB-2000;	2000US-0184800

XX  
XX  
/DTCT- ) DTCTCENMC

PA (YEDA ) YEDA RES & DEV CO LTD.  
XX

PI Bellenson J, Smith D, Lancet  
xy

DR WPL; 2001-290713/30.

PT New polynucleotides  
PT sensation for identifi-

XX  
PS Claim 8: Page 263: 1857nn: English

XX  
XX  
The present sequence is one of a n

The present sequence is one of a number of isolated polynucleotides which encode polypeptides involved in olfactory sensation. The polynucleotides can be used in screening for olfactory agonists and antagonists. The methods allow for the determination of primary scents and the identification of the odour receptors used to detect these primary scents. The methods also enable determination of secondary scents and the identification of combinations of odour receptors that are involved in detecting such secondary scents. This enables the construction of a scent representation (also called a scent fingerprint or scent profile), which may be used to re-create and edit scents. Libraries of olfactory receptors are useful for determining the interaction pattern of a composition with the receptors and can be used for determining differences in the olfactory faculties of different individuals.

\$Q\$ Sequence 936 BP; 222 A; 226 C; 188 G; 300 T; 0 other;

Query Match	42.5%	Score 400.4	DB 22	Length 936
Best Local Similarity	64.9%	Pred. No. 5.2e-113		
Matches 593; Conservative	0	Mismatches 321	Indels 0	Gaps 0

**Oy** 1 atgatactggatgaacagcaacaacatacagaagtattactgcgttggtttccaaatcga 60  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
**Dd** 1 atgatactgaaaaatgaagttcttcttcacacggaattaccctaactgggtttctcttaagaag 60  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
**Oy** 61 ccttgctgcaggagttccacactcctbtgtagcttctcttgattttcttaacactgtgacacattt 120  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
**Dd** 61 ccctacagctgagacagatcctctcttctgtgcttctttgacctctcatacatctcacattgcgt 120  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||





Db 493 gacatcgtgcatggtcagtggaatagaccaccctgtgacagtcacccctcacccctg 552  
OY 493 cagcgtccactctgagaccctatgtatagatcactctctgtgaaagtcctcagc 552  
Db 553 cagcgtccactctgagaccctatgtatagatcactctctgtgaaagtcctcagc 612  
OY 553 ctaagatcctctgtgttgaagacaacagcaaatgagcgtgaactatctctgtcagtgag 612  
Db 613 atcaagctgctgtgtgtgagacacagcaggttaacgagcgtgagcctttgtgtcagtaac 672  
OY 613 ccttcacatcaatcaccctgacacactcctatacatatgctttatgttgcagaga 672  
Db 673 ccttcacatcaatgctgtgtgagacacagcaggttaacgagcgtgagcctttgtgtcagtaac 732  
OY 673 gattgagagatcagctgtcgtgaaggtcagacaagaagcgttgagacatgtgtccact 732  
Db 733 ggttgagagatcagctgtcgtgaaggtcagacaagaagcgttgagacatgtgtccact 792  
OY 733 ctaattgtgtgtctctcttcttataagtaagcgtctgtgtgtacgtgacacactcgc 792  
Db 793 ctgacagctgtgtacacatcttcttataagtaagcgtctgtgtgtacgtgacacactcgc 852  
OY 793 cccagctccaaagacaaagaaagatgttctctctctctatcgtgaatcgtgcacccatg 852  
Db 853 agtagatccagagaccagagcaagttgttctctctctctacacgtgtgttaaccgcagatg 912  
OY 853 cgtgaatccctataatatacacttaggaagaaggtgaagaagcgtttaaaggtgtg 912  
Db 913 cttaacccctctatattatccttgaaggtatcaagaaggtgaagaagcgtttaaaggtgtg 972  
OY 913 gttgcaagagctctcttcaatcaagaataa 942  
Db 973 ctgacaaagcctctgaggaataatattta 1002

## RESULT 11

AAH32306

AAH32306 standard: DNA: 1011 BP.

AAH32306:

30-JUL-2001 (first entry)

Human olfactory receptor polynucleotide, SEQ ID NO: 879.

Human: olfactory receptor; OR: primary scent determination;  
secondary scent determination; polypeptide library; odour receptor;  
scent profile; scent fingerprint; scent representation; ds.

Homo sapiens.

MO200127158-A2.

19-APR-2001.

06-OCT-2000; 2000MO-US27582.

08-OCT-1999; 99US-0158615.

24-FEB-2000; 2000US-0184809.

(YEDA ) YEDA RES &amp; DEV CO LTD.

(DIGI-) DIGISCENTS.

Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

WPI; 2001-290713/30.

New polynucleotides which encode polypeptides involved in olfactory

sensation for identifying olfactory agonists and antagonists -

Claim 8; Page 531; 1857Pp; English.

CC The present sequence is one of a number of isolated polynucleotides  
CC which encode polypeptides involved in olfactory sensation. The  
CC polynucleotides can be used in screening for olfactory agonists and  
CC antagonists. The methods allow for the determination of primary  
CC scents and the identification of the odour receptors used to detect  
CC these primary scents. The methods also enable determination of  
CC secondary scents and the identification of combinations of odour  
CC receptors that are involved in detecting such secondary scents.  
CC This enables the construction of a scent representation (also called  
CC a scent fingerprint or scent profile), which may be used to re-create  
CC and edit scents. Libraries of olfactory receptors are useful for  
CC determining the interaction pattern of a composition with the receptors,  
CC and can be used for determining differences in the olfactory faculties  
CC of different individuals.  
XX  
SQ Sequence 1011 BP; 217 A; 252 C; 202 G; 340 T; 0 other;

Query Match 38.7%; Score 364.2; DB 22; Length 1011;

Best Local Similarity 62.9%; Pred. No. 7.3e-102;

Matches 564; Conservative 0; Mismatches 333; Indels 0; Gaps 0;

OY 12 aaatcagcatcacacaggaattatctctgtgttctcaagtcagcttgcgtga 71  
Db 9 aaatcaggttttgaagacttcttcttacttctgtgatttcttaactgtccatctga 68  
OY 72 gttccacactgtgtgtctcttcttcttcttcttcttcttcttcttcttcttcttctt 131  
Db 69 agtagtctcttctgtgttacttcttcttcttcttcttcttcttcttcttcttcttct 128  
OY 132 catattctagtgtacagcgttgaacacaaactatatacccatgtatttttcttctac 191  
Db 129 catcatcctcgtcatatccctgactccatctccacactcccatgtactcttctcttc 188  
OY 192 caattctacactccctggtatcttgttaacacacatgtacagtcacaaagtctagtaa 251  
Db 189 aaatctctatcttctgtacttctgttaccacacagcttcttcttcttcttcttctt 248  
OY 252 ttatgcagcatcagaaagtaacatgatctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 311  
Db 249 tctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 308  
OY 312 tctgtgcttgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 371  
Db 309 tctgtgactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 368  
OY 372 agctattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 431  
Db 369 agctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 428  
OY 432 ggcagccgcatcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 491  
Db 429 ggcgtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 488  
OY 492 ccaagctcactctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 551  
Db 489 ctgtgatacccatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 548  
OY 552 gctcaagatctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 611  
Db 549 tctgtgattatcatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 608  
OY 612 gcttctcatcctaataccctgagactctatcttatacatatgctttattgttcagagc 671  
Db 609 catcttcttctcatcactcctcatcctcctcctcctcctcctcctcctcctcctcct 668  
OY 672 agtatgagatagacagctgtcgtgaaggtcagacaagaacatctgtgagatctgtgtcca 731  
Db 669 tgtactgtgagcatgacataacacagcgtgtcagaaagtgtcttagaagatgtgtgagcca 728  
OY 732 tctaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 791  
Db 729 tcttatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 788







```
Db 61 cccagctgagatgctcttttataagcaltctctctatttctgacccactt 120
QY 121 ggcattctgacattatctatgtacagcctggaaaccaacttatacccccattat 180
Db 121 gggaaactcaaccatcatcttcttcccgctggagccggctccatcacccatgtac 180
QY 181 ttttcttaccatctcatctccctggatctcttcttaccacacatgtacgtccacaa 240
Db 181 tttcttccatgaacacctctccctcttggacctgtcttgcctactagtccaatccccaa 240
QY 241 atgctagtaaatltatgcaagcatcaggaagaatcaatcagttacgtgctgtgagccag 300
Db 241 atgctagtaaatltatgcaagcatcaggaagaatcaatcagttacgtgctgtgagccag 300
QY 301 ctcttcatacttctgaccttgggacctgactgaatactctctctgacgctacgtctt 360
Db 301 ctcttcatacttctgaccttgggacctgactgaatactctctctgacgctacgtctt 360
QY 361 gataggttctgactatttctgacctctccatctcaatctcaatctcaatctcaatctc 420
Db 361 gacccgtaagtgtgacgtgtgctggccgccccctccgctacacccgcatcagaaacccagctc 420
QY 421 tgcctccagtttgcagccgcatccctgggttactgtgtttagtaactcagttgtgtct 480
Db 421 tgcctccagtttgcagccgcatccctgggttactgtgtttagtaactcagttgtgtct 480
QY 481 accctgactctccagctgacctctgtgacccctatgtgataagatacttctctgtgaa 540
Db 481 accctgactctccagctgacctctgtgacccctatgtgataagatacttctctgtgaa 540
QY 541 gtcctgcaactctcaagtctctgtgtgtgtgagacaacagcaaatgaggtgaactatc 600
Db 541 gtcctgcaactctcaagtctctgtgtgtgtgagacaacagcaaatgaggtgaactatc 600
QY 601 ctgtcagtgagctctctccatctaaaccccgacacatccatctatacatatgtcttt 660
Db 601 ggtgtctgcaactctctctcaactgacgtccacgaatcagatcgtgtctctcaactgctc 660
QY 661 atgtccgagcagatattgaggaatacagctgtctgaaggtcgaacaaagaatttgggaca 720
Db 661 atgtccgagcagatattgaggaatacagctgtctgaaggtcgaacaaagaatttgggaca 720
QY 721 tgtgttcccatctaatgtgtgtgtctcttcttataagacagcgtctctgttactgt 780
Db 721 tgtgttcccatctaatgtgtgtgtctcttcttataagacagcgtctctgttactgt 780
QY 781 caacacacttgcacagctcacaagcaagcaaggaagatgttctctctctatagtatc 840
Db 781 ctccgccaagaacagcaagcaagcaaggaagcaaggttccctgtctactcgttg 840
QY 841 atgcaaccatgtcgaatctcccttatataatacacttagaacaagaaggtgaagaagc 900
Db 841 atgcaaccatgtcgaatctcccttatataatacacttagaacaagaaggtgaagaagc 900
QY 901 tttaaaggttgttgcagaag 922
Db 901 ctgaggaaggttctgtgggaaag 922
```

Search completed: February 26, 2002, 05:05:19  
Job time: 3117 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 26, 2002, 03:41:21 ; Search time 45.65 Seconds

(without alignments)  
4673.434 Million cell updates/sec

Title: US-09-755-017-1

Perfect score: 942

Sequence: 1 atgaattgggttaatgacag.....tcttcttaatcaagaataaa 942

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCrUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	216.8	23.0	1290	2	US-08-827-291A-1
2	211.6	22.5	966	3	US-08-748-506-7
3	211	22.4	966	3	US-08-748-506-8
4	200.8	21.3	1713	2	US-08-467-948A-1
5	200.8	21.3	1713	3	US-08-467-947A-1
6	200	21.2	900	4	US-09-085-371-5
7	197.2	20.9	966	3	US-08-748-506-5
8	190.4	20.2	966	3	US-08-748-506-6
9	183.6	19.5	984	3	US-08-748-506-9
10	118.4	12.6	1474	1	US-08-465-980-1
11	118.4	12.6	1474	2	US-09-053-303-1
12	118.4	12.6	1474	5	PCT-US95-07093-1
13	94.4	10.0	1828	3	US-08-988-876-2
14	90.2	9.6	1320	1	US-08-599-252-84
15	90.2	9.6	1320	1	US-08-436-074-57
16	90.2	9.6	1320	5	PCT-US96-06352-84
17	90.2	9.6	1320	5	PCT-US96-06353-84
18	57.8	6.1	7218	5	US-08-232-463-14
19	41.6	4.4	1260	3	US-09-082-088-1
20	41.4	4.4	1260	1	US-08-599-252-83
21	41.4	4.4	1260	1	US-08-436-074-56
22	41.4	4.4	1260	5	PCT-US96-06352-83
23	41.4	4.4	1260	5	PCT-US96-06353-83
24	39.8	4.2	2012	3	US-08-706-281A-7
25	39.8	4.2	2012	4	US-09-097-231-7
26	36.6	3.9	1015	1	US-08-671-525B-3
27	36.6	3.9	1015	1	US-08-672-109B-3

28	36.6	3.9	1015	1	US-08-842-045-3	Sequence 3, Appl1
29	36.6	3.9	1015	2	US-08-842-238-3	Sequence 3, Appl1
30	36.6	3.9	1015	3	US-08-629-338B-3	Sequence 3, Appl1
31	36.6	3.9	2012	1	US-07-866-560-3	Sequence 3, Appl1
32	36.6	3.9	2012	1	US-08-077-673-3	Sequence 3, Appl1
33	36.6	3.9	2012	1	US-08-478-992-3	Sequence 3, Appl1
34	36.6	3.9	2012	1	US-09-105-298-3	Sequence 3, Appl1
35	34	3.6	289	4	US-09-007-005-17	Sequence 17, Appl1
36	34	3.6	289	4	US-09-244-796-17	Sequence 17, Appl1
37	32.4	3.4	753	4	US-08-998-416-574	Sequence 574, App
38	31.8	3.4	1101	3	US-08-945-056-5	Sequence 5, Appl1
39	30.4	3.2	603	3	US-08-609-334-10	Sequence 10, Appl1
40	30.4	3.2	603	3	US-08-609-334-15	Sequence 15, Appl1
41	30.4	3.2	1092	4	US-09-077-675A-15	Sequence 15, Appl1
42	30.4	3.2	3129	4	US-08-077-675A-14	Sequence 14, Appl1
43	30.2	3.2	1101	3	US-08-945-056-7	Sequence 7, Appl1
44	30.2	3.2	80246	4	US-09-078-294-4	Sequence 4, Appl1
45	30.2	3.2	80595	4	US-09-078-294-3	Sequence 3, Appl1

#### ALIGNMENTS

RESULT 1

US-08-827-291A-1

Sequence 1, Application US/08827291A

Patent No. 5874243

GENERAL INFORMATION:

APPLICANT: Macina, Roberto

APPLICANT: Sathe, Ganesh

TITLE OF INVENTION: NOVEL OLRCCL5 RECEPTOR

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY:

ZIP: 19406

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/827,291A

FILING DATE: 28-MAR-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: King, William T

REGISTRATION NUMBER: 30,954

REFERENCE/DOCKET NUMBER: GP50001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-5015

TELEX: 610-270-5090

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1290 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-827-291A-1

Query Match 23.0%; Score 216.8; DB 2; Length 1290;  
Best Local Similarity 52.2%; Pred. No. 4.7e-60;  
Matches 482; Conservative 0; Mismatches 442; Indels 0; Gaps 0;

OY	1	atgaatttgggaaattgcaacacacacaaaggaatttattcttgcgttgattcttcagatga	60
Db	296	ATGGCAAGGAGGAATTTGGACCTTCMACTCGACTTATCTTCCGTGGAAATCTTCATATAC	355
OY	61	ccttgatcgaggattccacactccttggtgtctcttcttatttcttaacactgtgacacttt	120
Db	356	AGCCCCACCCACACTCTTCCTTCTTCTTCTGCTGGCATCTTTTCAATGGCCTTTCATG	415
OY	121	ggcaactctgaacattattctcagtgtaacgcgcctgtgaacccaacttcaataccccatgat	180
Db	416	GGAACATCTGTCAATGGTGTCTCCATCTACCTGGACACCCACTCCACACCCCACTATAC	475
OY	181	ttttcttccaaactatatacactccctggaaactctgtttaacacaatgtaacagttccacaa	240
Db	476	CTCCCTCCTCAACCAACTGCTCTCTCATATGACCTTATGCTATATGCAACACCTATCCAA	535
OY	241	atgctagtaaatatattatgcaacatcagaagaaagtatcaattacgtttacgtgtgtgaacag	300
Db	536	ATGGCGCTAACATACCTACGCTGTGGCAGAGATCCATTTCATATGGCTGTGTGGCACACAA	595
OY	301	ctttcaatattcttgaccttgagggtggaactgaatactctccctggccgtcaatgctctt	360
Db	596	ATTTCCTTCATATACATACACGCTTGGCTGTGATCTTCCTTTTGGCTGTATGCTTAT	655
OY	361	gataagtttgtagtatttttggtgcctctccatctactcaatgatatatgatgacaagagact	420
Db	656	GACCGCTACACTGCCATTTTCCACCCCTCTAAGATATACCAATCTCATGAGCCCTTAAAT	715
OY	421	tgcctcagttgagcagcgacatcccttggtatacgtgttttagtaactcaatggtgtgtct	480
Db	716	TGTGGACTTATGATGCTCTTCTTCTGGATCTGAGTCTACAGATGGAAATCATTTATGCT	775
OY	481	accctgaactcgaagctgcaactctgtgaacccctatgataagataactactctctcgtgaa	540
Db	776	GTAACCAATTTTCTTCCTCTCTACTGTGGGTCTCGGGAATAGCCACACTTCTTCTGTAG	835
OY	541	gtccctcactctgctcagaattatttttgtttggaagacagcaatgagagcttgaactatct	600
Db	836	TTACCTTCTCCTACTAATCTCTCTATGCAATGACATCTCAATTTTGAAAAGGTTATTTTTC	895
OY	601	cttgacagtgagctctccatacctaaccctcgacactatctctatacatgatctctt	660
Db	896	ATTTCGTCTATAGTAATGCTGTGTTTTCCTGTTTCCGTTGCATCATATGCTTCCATGCTGA	955
OY	661	atgttcagagcagtatatgagatatcacgtctgtcgtgaagtcgaacaaagcatttggaca	720
Db	956	GTTATTCCTGCTGTCATTTCAATGAGGATGTGGAGAGGCTGTCGCAAAAGCTTTCACGACC	1015
OY	721	tgttgctccatctaaatttgggtgtctcttttataagtaagccogtctctgtgtaactgt	780
Db	1016	TGTTCTCTCTACCACTGAGTGGTGGAAATGCTTATAGGAGACGGTTTGTTCTATGATTA	1075
OY	781	caaccacacttcgcccacactcgaagacgaacgaagaatgtgttctctctctcatgaaatc	840
Db	1076	CAGGCCACATCTGATATGCTCTCCCAACAGCAAGAAACCTGGTGTCTGTATTTACACATTC	1135
OY	841	atgacaccaatgctgaaatccctctatatacatcaattaggaacaagaggtlaaaggaagc	900
Db	1136	CTCACTCCCATGCTGAATCCCTCATCTACATGAGCTCCGACACAGAAGATGACAGACGA	1195
OY	901	tttcaaaagtttggttgcaagatc	924
Db	1196	TTTCATGAAGATCTCAGGAAAGGCC	1219

RESULT 2  
US-08-748-506-7  
: Sequence 7, Application US/08748506  
: Patent No. 6159707  
: GENERAL INFORMATION:  
: APPLICANT: Romett et al.  
: TITLE OF INVENTION: NOVEL SPERM RECEPTORS  
: NUMBER OF SEQUENCES: 31

1 CORRESPONDENCE ADDRESS:  
2 ADDRESSEE: Leydig, Volt & Mayer, Ltd.  
3 STREET: Two Prudential Plaza, Suite 4900  
4 CITY: Chicago  
5 STATE: IL  
6 COUNTRY: US  
7 ZIP: 60601-6780  
8 COMPUTER READABLE FORM:  
9 MEDIUM TYPE: Floppy disk  
10 COMPUTER: IBM PC compatible  
11 OPERATING SYSTEM: PC-DOS/MS-DOS  
12 SOFTWARE: Patentln Release #1.0, Version #1.25  
13 CURRENT APPLICATION DATA:  
14 APPLICATION NUMBER: US/08/748,506  
15 FILING DATE: 08-NOV-1996  
16 CLASSIFICATION: 435  
17 PRIOR APPLICATION DATA:  
18 APPLICATION NUMBER: US 60/033,751  
19 FILING DATE: 09-NOV-1995  
20 CLASSIFICATION: 435  
21 ATTORNEY/AGENT INFORMATION:  
22 REFERENCE/DOCKET NUMBER: 74940  
23 TELECOMMUNICATION INFORMATION:  
24 TELEPHONE: 312-616-5600  
25 TELEFAX: 312-616-5700  
26 INFORMATION FOR SEQ ID NO: 7:  
27 SEQUENCE CHARACTERISTICS:  
28 LENGTH: 966 base pairs  
29 TYPE: nucleic acid  
30 STRANDEDNESS: double  
31 TOPOLOGY: linear  
32 MOLECULE TYPE: DNA (genomic)  
33 US-08-748-506-7

Query Match	22.5%	Score 211.6	DB 3	Length 966
Best Local Similarity	55.2%	Pred. No. 1,9e-58		
Matches 412	Conservative 0	Mismatches 334	Indels 0	Gaps 0
QY	163	ctcaataccccaatgattttttcttaccacaatcatcaactcctcgtgactttgttaacac	222	
Db	181	CTACACACCCCCAGTACTCTTCTGTGGCCAACTGTCTCTCCGAGAAATGGCATACT	240	
QY	223	acatftacagtcgcccaatgctcgtatgaattatgatgcagcatcaggaagaatcaagttat	282	
Db	241	TGCTCTGTCAATACCCAGAAATGCTGCAGAGCCTGTGATGTAGGCCCCAGGGATCTCTTGG	300	
QY	283	cgttgagctgtgaagccagctttccatatattctcgccttcgggggactacgaatattcttc	342	
Db	301	GAGGTTGTGGCTCACAGATGTCTCTTTCATATTCCTTGGATATACGTAGTGTGCCTTA	360	
QY	343	ctgagccgtcaatgctcttggatgaagtgttgtaactatttgcgcctctccattactcagtt	402	
Db	361	TTTGGAGCCAAATGAGCCTTTTGAGCCGCTATATATGCTTATATGTTCCCACTACCTATATGCAACC	420	
QY	403	atcatgaccacagagactcctgcctcagttgtgcacgcagcatcctggttactcgttttagt	462	
Db	421	CGAATGATCTCTGGGGATATGTGCCATTTTGGCAATTTGTCTATATGGGTATATGGATGATA	480	
QY	463	aactcagttgagttgttctaacccctgactctcagactgcacatcgtgacacccatagtatga	522	
Db	481	GTAGTCTGGGACAGACCAATTTATTTTCTCTTGAACCTCTGTGGACCCCTGTGAGATA	540	
QY	523	gataccttctcttgaagtcctcgcgcactgttcaagtattcttgttgaagaacaagca	582	
Db	541	GACCACTTCTTGTGAGACTTTCACACTCTCTCTGGCACTTGCCTGTGGTATGATACATCCAA	600	
QY	583	aatgagcttgaactatctcttgcagtgagctcttccatcaataacccttcaactcatc	642	
Db	601	AATGAGCTGCACATCTTTGTGGACAGACAGTCTCTGCATATTTAATGCCATTTTAACTGATC	660	
QY	643	cttatatacatgcttattatgtccgagcagattatgagatatacagcttgcctgaagtgta	702	



Db 661 ATTCTTCCATGTCAGAAATTCGTGCAAGTGTGATGCTTACCTGAGGGGCGC 720  
QY 703 caaaagcatttggagacatgtgttcccatcattatgtgtctcttttattagaa 762  
Db 721 CATAAAGTCTCTCTACTGTCATCTCAGCTACTTGTAGTACACACTTCTATGGCTCA 780  
QY 763 ggcgtctctgttactgtcaacacaccttgcgcagctcccaagaaagaaatgtgt 822  
Db 781 ACATCTGCCACCTATTGTAGAGTCCAGTGTGACCATCAGGAGAGTGGACAACCTCTTG 840  
QY 823 tctctctatgaaatcattgtcccatgtgtatcccttatataataacttagaac 882  
Db 841 GCCCTCTTATACATGATGATCATCATCTGATCCATCATCATACAGCTTAAGAAC 900  
QY 883 aagaggtaaagaaagccttaaaag 908  
Db 901 AAGGAGTAAAGGCTGCATGAGAG 926

RESULT 3  
US-08-748-506-8  
Sequence 8, Application US/08748506  
Patent No. 6159707  
GENERAL INFORMATION:  
APPLICANT: Romelt et al.  
TITLE OF INVENTION: NOVEL SPERM RECEPTORS  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leydig, Volt & Mayer, Ltd.  
STREET: Two Prudential Plaza, Suite 4900  
CITY: Chicago  
STATE: IL  
COUNTRY: US  
ZIP: 60601-6780  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/748,506  
FILING DATE: 08-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/033,751  
FILING DATE: 09-NOV-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
REFERENCE/DOCKET NUMBER: 74940  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-616-5600  
TELEFAX: 312-616-5700  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 966 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-748-506-8

Query Match 22.4% Score 211; DB 3; Length 966;  
Best Local Similarity 55.0%; Pred. No. 3e-58;  
Matches 415; Conservative 0; Mismatches 340; Indels 0; Gaps 0;

QY 163 cttaacacccacatgtatttttcttcaacatatacaaccctgtgacttgttaacc 222  
Db 181 CTACACACCCCGATGACTTCTTCTGCGCAACTGTCTCTCCGGAATATGGCATACT 240  
QY 223 acatgtacagtcacacaaatgtctatgaattatgtcagcagcatcagaaatgaatcag 282  
Db 241 TGCTCTGTCAATCAAGATGCTGACAGAGTGTGTGAGTGAGCCCGAGAGATCTTTCAG 300

QY 283 cgtgctgtgtagccacgttcttcatattctgtgacctggtgagcactgaatccttc 342  
Db 301 GTGGAGATGTGCACACAGATGTTTCTTCAATTTCTTGTATACAGATGCTGCTCA 360  
QY 343 ctgagcgtacgtcttcttgaatgtgtatgtatgttgcgcctctccatcactcagtt 402  
Db 361 TTGGCAGCCAGTGGCTTGAACCGCTATATGCTATATGTCCCACTCACTATGACAC 420  
QY 403 atcagtcacagagacgtctgctccagtgagcagcagcagcagcagcagcagcagc 462  
Db 421 CGAATGAGCTGTGAGATGTGCCCACCTTGGCAATGTTTATAGGATGAGATGAGATCA 480  
QY 463 aacacagtggtgtgtacacaccttgcacgtcagcagcagcagcagcagcagcagcagc 522  
Db 481 CTAGTCTGGGACACAGACCAATTTATTTTCTTCTTGAACCTCTGTGAGACCTGTGAGATA 540  
QY 523 gataccttctctgtgaagtcctcagcagcagcagcagcagcagcagcagcagcagcagc 582  
Db 541 GACCACCTCTCTGTGATCTTCCACCTCTCCGACACTTGGCTGTGATATACATCCCA 600  
QY 583 aatgagctgacatcttctgtcagtgagcttctcactaaacccctgaactcactc 642  
Db 601 ATTAGGCTGCTATCTTGTGTGATGTGCTCTGCAATATCTAGCCCTTTTGTCTGATC 660  
QY 643 ctatacatatgcttcttctgtccagcagcagcagcagcagcagcagcagcagcagcagc 702  
Db 661 ATTATTTTATGTCAGAAATTTCTGTCAGTGTGATGATGCTTCACTGAGGGGCGC 720  
QY 703 caaaagcatttggagacatgtgttcccatcattatgtgtgtctcttcttattagaa 762  
Db 721 CACAAAGCCCTTCAACCTGTCCCTCCACCTACTTGTAGTACACTCTTTTATGAGCTCA 780  
QY 763 ggcgtctctgttactgtcaacacaccttgcgcagcagcagcagcagcagcagcagcagc 822  
Db 781 GGATCTGTACTATTGTAGGCTTAAGCTAGCCACTCAGCAGAAATGACAAATCTCTG 840  
QY 823 tctctcttcttgaatcattgtcaccacagcagcagcagcagcagcagcagcagcagcagc 882  
Db 841 GCCCTCTTCTACACAGCATGATCATCTGTTGAACCTTATCATCTATATGTTAAGGAC 900  
QY 883 aagaggtaaagaaagcctttaaaggtgtgtgc 917  
Db 901 AAGGATGTCAAGGACGACACTGAGAAATTTCTGCGC 935

RESULT 4  
US-08-467-948A-1  
Sequence 1, Application US/08467948A  
Patent No. 5998164  
GENERAL INFORMATION:  
APPLICANT: LI, YI  
APPLICANT: CAO, LIANG  
APPLICANT: NI, JIAN  
APPLICANT: GENTZ, REINER  
APPLICANT: BULT, CAROL J.  
APPLICANT: SUTTON III, GRANGER G.  
APPLICANT: ROSEN, CRAIG A.  
TITLE OF INVENTION: Polynucleotides Encoding Human G-Protein  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
STREET: 1100 NEW YORK AVE., NW, SUITE 600  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PATENTIN RELEASE #1.0, VERSION #1.30



Query Match 21.3%; Score 200.8; DB 3; Length 1713;  
 Best Local Similarity 52.4%; Pred. No. 7.9e-55;  
 Matches 442; Conservative 0; Mismatches 402; Indels 0; Gaps 0;

12 aaatgaagacatcatcaagaagattatctcgtcgtgttccagatcgacctgtcga 71  
 136 AAATCAACACATGCTCAGAGGTTCCCTACTGGGATTTCTCTGGGCCCAAGATTCA 195  
 72 gttccacctgtgtgtctctgtatcttcttaacatgagccatttggcaatcgac 131  
 196 GATCTCTCTTTTGGCTCTTCTCTCCCTGTTCTATGCTTCACTGCTGGGAGGAC 255  
 132 catattctagtgcacgcctgcgcacacacatccatccatgattttttcttac 191  
 256 CATCTTGGGGGTATCTACAGGACTCCAGCTCCACACCCCAATGACTTCTCTCTC 315  
 192 caatctacatcctcgtgactctgttttaacacacatgtaaccccaaatgtagtaa 251  
 316 ACACCTGGCGCTGCTCAACATCGCCTAGCTGCAACACAGTCCAGATGCTGTA 375  
 252 ttatgcagcatcagaagaatcaatcatctcgtcgtgtgtgtaacagctttcata 311  
 376 CTTCTTCATATCCAGCCAGCCCATCTCTTGTGCTGTTGCACTAGACTTCTCTT 435  
 312 tctgccttgggagtaactgaatattctcctcgtgcgtcactgcttctgtagttgt 371  
 436 TTTAGTTTGGACATACATGATGCTCTGTTGCTGATGCTGATGCTGATGCTGAT 495  
 372 agctatttgcgcctcctcaatcaatgataatcagtaacacagagactcgtccagt 431  
 496 GGCCATGCGCCACCTCTCCATATTTTCATCATGATGATGATGATGATGATGAT 555  
 432 ggcgcgcgcacccgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 491  
 556 GGGCATATCTTCTGGAACATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 615  
 492 ccagctgcacactcgtgaacctatgtgatagatcaacttctcgtgaagccctgact 551  
 616 AAGACTGCTCTTGTGGGCTCTGTAATTAACCACTTCTTGTGTAATCTGCTGCT 675  
 552 gctcaatlatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 611  
 676 CCTCAGCTGCGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 735  
 612 gctttccatcaataacccctgcacatcctatctatcaatgcttttatgtccagc 671  
 736 GTTATCTCTGCTGGACCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 795  
 672 agatagaagatacagctcgtgaagtgcaaaaagacatttggagatgtgttccca 731  
 796 CATCTTGAAGATCCAGTCTGGGAGGCGCGAAGAGCCCTTCTCCACCTGCTCTCCA 855  
 732 tctaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 791  
 856 CCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 915  
 792 gccacagctccaagacaaagaagatgttctctcttcatggaatcatgaccat 851  
 916 CCGCCATCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 975  
 852 gctg 855  
 976 GATG 979

RESULT 6  
 US-09-085-371-5  
 ; Sequence 5, Application US/09085371  
 ; Patent No. 6218358  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Firestein, Stuart  
 ; APPLICANT: Zhao, Haiqing  
 ; TITLE OF INVENTION: Functional Expression of, and Assay for, Functional Cellular Re

TITLE OF INVENTION: Vivo  
 FILE REFERENCE: P01511US2 / 09805059  
 CURRENT APPLICATION NUMBER: US/09/085.371  
 CURRENT FILING DATE: 1998-05-19  
 PRIOR APPLICATION NUMBER: US 08/891,243  
 PRIOR FILING DATE: 1997-07-10  
 PRIOR APPLICATION NUMBER: US 60/045,961  
 PRIOR FILING DATE: 1997-05-07  
 NUMBER OF SEQ ID NOS: 6  
 SOFTWARE: PatentIn version 3.0  
 SEQ ID NO 5  
 LENGTH: 900  
 TYPE: DNA  
 ORGANISM: Rat  
 US-09-085-371-5

Query Match 21.2%; Score 200; DB 4; Length 900;  
 Best Local Similarity 53.0%; Pred. No. 1e-54;  
 Matches 459; Conservative 0; Mismatches 395; Indels 12; Gaps 1;

31 gagttatctcgtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 90  
 34 gaattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 93  
 91 tttctgttttctcactgtagcatcttcttgcaatctgacatattatctgtgtcagc 150  
 94 tctctctgagactatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 153  
 151 ctgagacacacatcatalaccccatgattttttcttaccatctacacatcctgagat 210  
 154 aaccacccacacccacacacacacacacacacacacacacacacacacacacacac 213  
 211 ctgtgtacacacacacacacacacacacacacacacacacacacacacacacacac 265  
 214 attgtgtatgcacgtatcatctcctcctcctcctcctcctcctcctcctcctcctc 273  
 266 -----ggaagatcatagttatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 318  
 274 aacatgagacagctgtatctccttctgtgaggaacacacacacacacacacacacacac 333  
 319 ttgggggtactgaatactctcctcgtgcgacatgctctgtatgatttgaagctatc 378  
 334 ttgggttcacagagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 393  
 379 ttgcgcctctcactatcagttatcaatgacacagagactcgtccacagtgagcagc 438  
 394 tgcataccactcactacacccgctcatgtgcagtagccgctatgtgtcagatgcagct 453  
 439 gcatcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 498  
 454 ggaatcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 513  
 499 ccactctgtgacccctatgtagatcatcttctctgtgaagtcctgcacatgcaag 558  
 514 tcttaactgtggtgcacacacacacacacacacacacacacacacacacacacacac 573  
 559 ttatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 618  
 574 ctgtcatgcatgacatgtccacagcagagacttcaagacttctgtcctgtgcattttat 633  
 619 catcaataacccctgacacatctccttataatcatgttttatgttcgagagattgt 678  
 634 ctgtgtgagacccgtctcgtgtacatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 693  
 679 aggatacagctgtcgtgaagtgcaaaaagacatttggagatgtgtgtccatcaatt 738  
 694 cgcataccctcagctcgtgcgcgcacataaagcctttcaactcgtgcctccacactact 753  
 739 gtgtgtctctttttatagtaacagcgtctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 798  
 754 gtgtgtatcatcttatacagcagatatttcatctatgcagagcctaagacactctca 813



Tue Feb 26 09:22:07 2002

us-09-755-017-1.rni

Page 7

```

? REFERENCE/DOCKET NUMBER: 74940
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 312-616-5600
? TELEFAX: 312-616-5700
? INFORMATION FOR SEQ ID NO: 6:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 966 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? OS-08-748-506-6

```

Query Match	20.28;	Score 190.4;	DB 3;	Length 966;
Best Local Similarity	53.48;	Pred. No. 1.3e-51;		
Matches 398;	Conservative 0;	Mismatches 348;	Indels 0;	Gaps 0;

QY	163	cttataaccccaatgtaattttttttcttaacaaatcatcaatccctggatcttcttgtaacc	222
Db	181	ctacacaccccccatgtaattctttcttggtcccaacttgctctctcgtggattggtctaact	240
QY	223	acatgtaacagtcuccacaacaatgtaagtaattatagcagcatcagaagaatcaagttat	282
Db	241	tgctctgtcattatcccaagatgctccaaagccttgtagtgtagagccacaaagatcttagg	300
QY	283	catggtctgtgaagccagcttttcataattctgacctggggactactgaataatctctc	342
Db	301	gagggatgtgcacacacagatgtttttcttcacattttttggcattatcgatgctgccta	360
QY	343	ctggccctgaatgaccttggtagtttgtagatatttgctgacctccatctactaagtt	402
Db	361	ttggacaccacatggccttttgacccgtgcattggccattatgctccacacactatgtaacc	420
QY	403	atactgaccagaagactctgcctcaagtttgcaagcccatccatctgggtactgattagt	462
Db	421	cgatgatgctgtgagatgatggtcccatatttggcattttttcatgggaatgggatgata	480
QY	463	aactcaatgltgltgtctactaacctgactctccaagctgccaactctgtgaacctatgtga	522
Db	481	gtagctgtgggacagaccatttttatatttttcccttgaaactttctgtgacctgttgagata	540
QY	523	gataactttctctgtgtaagatccctgcacatgctcaagtatcctgtgttggaacaacga	582
Db	541	gaccacttcttctgttgacatttccacacttctcttgacacttctctgtgtgataatccaa	600
QY	583	aatgaagctgaactatccctgtgacagtgagctctccatctaataaccctggacaatc	642
Db	601	aacgagctgccattttgttgacagcaaatctctgtatattctagttccatttttgggtatc	660
QY	643	ctataatcatatgcttttattatgtgcagcagatattgagaatacagctgtcgaagtcga	702
Db	661	ctttattcttattgacaaatattctgttgacagtgctggtgaaagccttaacctgaagggcgc	720
QY	703	caaaaagcaatttgggacaatgtgtgttcccatctaaatgtgtgtgtccttttattagtaa	762
Db	721	catataacctttttccactgtttcttcacaccttactttatgacacactttttatggctct	780
QY	763	gcgctcctgtgtaaccgcgaacacaccttggcccaagctccaagaacgaagaatggtt	822
Db	781	gtgttccttttaccttattttgagggccccaagtctatgaccattccacaggaatggacaaactcttg	840
QY	823	tctctctcttaagtaaatcatgtgcccacatgtgatacccttatalataaaactgaagaac	882
Db	841	gccctctttctacacagcagatgacattccatgtgtgaacctttacattctacagttaagaaac	900
QY	883	aaggaggtaaagaagaccttaaaag	908
Db	901	aaggaaagtcaaggcagcactgagaaag	926

RESULT 9  
US-08-748-506-9

Sequence 9, Application US/08748506  
Patent No. 6159707

GENERAL INFORMATION:  
APPLICANT: Ronnett et al.  
TITLE OF INVENTION: NOVEL SPERM RECEPTORS  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leydig, Voit & Mayer, Ltd.  
STREET: Two Prudential Plaza, Suite 4900  
CITY: Chicago, Illinois 60601

COUNTRY: US

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compati

SOFTWARE: Patentin Release #1.0, Version #1.25

APPLICATION NUMBER: US/08/748,506

CLASSIFICATION: 435

APPLICATION NUMBER: US 60/033,751

CLASSIFICATION: 435

REFERENCE/DOCKET NUMBER: 74940

TELEPHONE: 312-616-56  
TELEFAX: 312-616-5700

SEQUENCE CHARACTERISTICS:

LENGTH: 984 base pairs  
TYPE: nucleic acid

STRANDEDNESS: ac

MOLECULE TYPE: DNA (genomic)  
8-748-506-9

Query Match 19.5%; Score 183.6; DB 3; Length 984;

```
Matches 444; Conservative 0; Mismatches 434; Indels 0; Gaps . 0;
```

Qy	13	aagacagcaatcaaacaggagttatctctcgtgttctccagatcgtacttgcgtggag	72
Dy	13		72
Dy	31	AATGGGACTTTGGCTCGGAGTTTCATCTTTGAGGGGTACCTGTGGCGGACACTGAA	90
Qy	73	tttcacaccttgcgtctcttgatcttcaactctgacacatcttgacaatcagac	132
Dy	73		132
Dy	91	ATCCTCTTTTCCACTGCACCTTCTGGCTACTTGGCTTCCTCATGGGGCAATGCTC	150
Qy	133	atattctagtctacgcctctgagacacaatcatacccccatgatatctttcttacc	192
Dy	133		192
Dy	151	ATAATTACCATCACCTGTGTGAGACACCGACTGCAGACGCCATGTACTTTTTCAGC	210
Qy	193	aattatcatcctccgtgactcttgttatacacaacatgtagtcccaacaaatgcagtaat	252
Dy	193		252
Dy	211	ACCTTCCTTTTGGAGGTGTTTATTAACATCTCTATCCGCCAGCTCTCACCATC	270
Qy	253	ttatgcagcatcagaagaatcaatcagttatcgtgcgtgttagccagctttcatactt	312
Dy	253		312
Dy	271	ATTCTGTCAGGAGGACAAAGATTCCCTTTGGGGTCTGCTTTCACAGGCTTCGTAT	330
Qy	313	ctggaccttgggagctactgataatctctccctgcgcgtcatgcttcttgatagttgta	372
Dy	313		372
Dy	331	CTTCTCGTGGGGCAACAGGTTTTTCTCTTTGGCTCGTTATCCGTGAGCCGTTTCTG	390
Qy	373	gcatcttgcgtgcctcccaatctcaatgatatcatatgacaagaagactctgccccagtg	432
Dy	373		432
Dy	391	GCCATCTGCAAACTCTTACATATTATCCACACATATGAGCCCAAGAGATGTCCTTCCTC	450
Dy	391		450

Oy	433	gaagcgcgacccctgggttaacggttttaagtaaaccaagtggtgtgtctacccggaactc	492
Db	451	gTTACTGCTGTTATTATTTTGGGCTTCCCTTCATAGGCCAGATCCAGTTGATGCTTTCC	510
Oy	493	cagctgcacactcgtlgaaacccctatgtagatgataccttctctgtgaagtcctcgacatg	552
Db	511	AAGACATTTTCTGTGGTCCAAACATTAATTCCTCACTTTTCTGTGATTTTGGAGCCCTG	570
Oy	553	ctcaagatcatcttctgtgtgagacacacgcaatbtaggctgaacatctcctgtgcagag	612
Db	571	GCAAATCTCTCTGTTTCAAGAAACAGGCTATTATAGATGCTGTTTTTAACTTCCTGCTGA	630
Oy	613	ctctcccatcctaatcaaccctgaacactcatcccttatcatcatatgctttatctgcagaca	672
Db	631	ATTGGCTTTTGGCTCTCTTTCTTATAGCCATCTTTGCATTAACAGCAATATAGTAGTCAC	690
Oy	673	gtattgagatagaagctcgtcgtgaagtcgcgacaagaagcaatctggacatgtgtgccat	732
Db	691	ATAGGAAACATCCCTTCAGCCAGGAGAGAGAGACCTTTTTCACCTCTCTCCAT	750
Oy	733	ctaattcgtggtctcctcttttataagacaagcgtctcgtgtaacctgcgaacacacctcg	792
Db	751	CTCATTTGCCCTCTCTCTATATATAGGCCAGCTGTGCATTTATATACCTGAAGCCAAAGAC	810
Oy	793	cccagctccaaagaccaaagaagaatggtctctctcttataagaaatcattgcacacatg	852
Db	811	AGAACACAGAGTGACACCAACAGAGAGGCTCTCTTGTAACAATGAGTTGACACCCCTT	870
Oy	853	ctgaatcccttatataataacacttaagaacaagaaggt	890
Db	871	CTGAACCTGTCAATCTACACCTCGGCAACAGCAGGT	908

RESULT 10  
US-08-465-980-1  
: Sequence 1, Application US/08465980

```

1  GENERAL INFORMATION:
2  APPLICANT: Soppet, Daniel R.
3  APPLICANT: 11, Y1
4  APPLICANT: Rosen, Craig A.
5  APPLICANT: Ruben, Steven M.
6  TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HPR470
7  NUMBER OF SEQUENCES: 8
8  CORRESPONDENCE ADDRESS:
9  ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI
10 ADDRESSER: STUART & OLSTEIN
11 STREET: 6 Becker Farm Road
12 CITY: Roseland
13 STATE: New Jersey
14 COUNTRY: USA
15 ZIP: 07068
16
17 COMPUTER READABLE FORM:
18 MEDIUM TYPE: Floppy disk
19 COMPUTER: IBM PC compatible
20 OPERATING SYSTEM: PC-DOS/MS-DOS
21 SOFTWARE: PatentL Release #1.0, Version #1.30
22 CURRENT APPLICATION DATA:
23 APPLICATION NUMBER: US/08/465, 980
24 FILING DATE: 06-JUN-1995
25 CLASSIFICATION: 536
26
27 ATTORNEY/AGENT INFORMATION:
28 NAME: Ferraro, Gregory D.
29 REGISTRATION NUMBER: 36,134
30 REFERENCE/DOCKET NUMBER: 328800-446
31 TELECOMMUNICATION INFORMATION:
32 TELEPHONE: 201-994-1700
33 TELEFAX: 201-994-1744
34
35 INFORMATION FOR SEQ ID NO: 1:
36 SEQUENCE CHARACTERISTICS:
37 LENGTH: 1474 base pairs
38 TYPE: nucleic acid
39 STRANDEDNESS: single
40

```

```

?      TOPOLOGY:  linear
?      MOLECULE TYPE:  CDNA
?      FEATURE:
?      NAME/KEY:  CDS
?      LOCATION:  274..1233
US-08-465-980-1

```

Query Match	12.68;	Score 118.4;	DB 1;	Length 1474;
Best Local Similarity	48.98;	Pred. No. 2.6e-28;		
Matches 317; Conservative	0;	Mismatches 331;	Indels 0;	Gaps 0;

QY	109	gtgacacattcttgcaatctgcacattatcttagtgcagccgcgcacaaattcat	168
Db	379	GTGGCAATGTTGGAAACTGATGTGTGTTTCAATCGTAAGACGAAACGACTTCAC	438
QY	169	accoccatgtaatttttcttaccaatctactcctgtaaccttgtaacocaaagt	228
Db	439	GCTCGGATGTCCTTTTCTTCCTCATCTTGACACCCTTAATGACGGCCTTATCCATCC	498
QY	229	acagtcaccaaatgctagtaaatatttgcagctcaagaaagtaacagttatcgtgyc	288
Db	499	ACCAATGCTTAAGATCTTGCCCTTTTCGTGTGATTCCGAGAGATTATACATTGAGCC	558
QY	289	tgtatgacccagctttcataattctctgaccttgaggactaagatactctccctggc	348
Db	559	TGTCTTACCAGATGTTCTTTATTCATGACCCCTTCACACATTGAATGCACATCTGCTG	618
QY	349	gtcatgtccctttagtaagtttgtatgataatttgcgcctctccatctacgtatcatg	408
Db	619	GCCATTGGCCTTTGACCGTTAATGTGGCCATTCGCCACCCACTGGCCCATGCTGAGTGCCTC	678
QY	409	caccaggaacccctgcaccagttgagacgcgcacatccctgggttaactgtttagtaactca	468
Db	679	AACATTAAGTAACAGCCCAATGTGGATCGCTGGCTGTGTCGGGATCCCTCTTTT	738
QY	469	gtgtgtgtgtctacccgaactcccaagctgcacactctgtgacccctatgtaatagcac	528
Db	739	TTCCACACTGCTCTGCTGATCAACGGGCTGGCCTTTCGCCATCTCATGTCTCTGCCAC	798
QY	529	ttctctctgtgaagtcctgtcaactgcagctcaagtatctctgtgttgagaacagaacaatgag	588
Db	799	TCTCATTTGTGCCACCAAGATGATGAAGTTGGCCATGCAGACACTTTGGCCCAATGTG	858
QY	589	gctgaacattctctgttcaagtagactcttccatctaatatcccccgcgaacatccattata	648
Db	859	GTAATATGTTTACTGCCATTCTCTCTGTCTATGGGCGTGGACGTAAATGTTCAATCTCTTG	918
QY	649	tcaatagcttttatttgcgcagacgatatgagaatacagctctgcgtgaagtgacaanaa	708
Db	919	TCTCATTTTTCGATTAATACGAACGGTTCCTGCACATGCCCTTCCAAGTACAGACGGGGCCAG	978
QY	709	gcatttggagatatgtgttccatctcaatagtgtgtctctcttttat	756
Db	979	GCCTTTGAACCTGTGTCAACAAATGTGTGTACTCCGCTTTAT	1026

RESULT 11  
US-09-053-303-1

1 GENERAL INFORMATION:  
2 APPLICANT: Soppet, Daniel R.  
3 APPLICANT: LI, Yi  
4 APPLICANT: Rosen, Craig A.  
5 APPLICANT: Ruben, Steven M.  
6 TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HPR470  
7 NUMBER OF SEQUENCES: 8  
8 CORRESPONDENCE ADDRESS:  
9 ADDRESSEE: CARELLA, BYRNE, BAIN, GILLILLAN, CECCHI  
10 ADDRESSEE: STUART & OLSTEIN  
11 STREET: 6 Becker Farm Road  
12 CITY: Roseland

STATE: New Jersey  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/053,303  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/465,980  
FILING DATE: 06-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferraro, Gregory D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-446  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1474 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 274..1233  
US-09-053-303-1

Query Match 12.6%; Score 118.4; DB 2; Length 1474;  
Best Local Similarity 48.9%; Pred. No. 2.6e-28;  
Matches 317; Conservative 0; Mismatches 331; Indels 0; Gaps 0;

QY 109 gtgacacatttggaatcgtgacattatctagtcgaagcctgagacacaaactcat 168  
DB 379 gtggcaatgtgtgaaactgcatgctgttcattcgttaagagacgaacgctgcac 438  
QY 169 accccatgattttttcttaccatctactcctgagcttctgtttacacaaatgt 228  
DB 439 gcttcgaatgaccttcttctgcatgcttgacacgaattgacgcttattccacatcc 498  
QY 229 acagtcacacaaatgctagtaaatltaagcagcatcagaaagtaacagttatcgtg 288  
DB 499 accatgacctgaagatccttgcccttttctggttattcccgagagattgacattgag 558  
QY 289 tctgtgacacagctttcatattctgagccttgaggagctactgaatctctcctg 348  
DB 559 tgccttaccagatgttcttattcattgaccccttcacacgaattgacacacatcctg 618  
QY 349 gtacatgcttggtagttgtagctattgttcgacctcctactactactgatactg 408  
DB 619 gccatggccttggacccttattgctgacacacacacacacacacacacacacacac 678  
QY 409 caccagagactcctcctcagctgagcagccgacccctgaggtactggttttagtaact 468  
DB 679 AACAAATACAGTAACAGCCAGATGGCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 738  
QY 469 gtgtgtgtctacacacacacacacacacacacacacacacacacacacacacac 528  
DB 739 TTCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 798  
QY 529 ttctctgtgaagtcctcctcagctgctcaagtatctgtgtgtgagacacacacacac 588  
DB 799 TCTATTGTGTCACCGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 858  
QY 589 gtgaacattctctcctcagctgagctctcctcactaataaccctgacacacacacac 648

DB 859 GTATATGCTTACTGCCATCTGCTGTCATGAGGCGGTGAGAGTAATGTCATCTCTTG 918  
QY 649 tcatatgcttattgtcgcagcagatltgagagatacagctcgtgaagtcgacaaaa 708  
DB 919 TCTATTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 978  
QY 709 gcatatgagacatgtgtcctcactaattgtgtgtcctcttcttatt 756  
DB 979 GCCTTTGAACCTGTGTGTACACATGCTGTGTGTGTGTGTGTGTGTGTGTGTAT 1026

## RESULT 12

PCT-US95-07093-1  
Sequence 1, Application PC/TUS9507093

## GENERAL INFORMATION:

APPLICANT: Soppet, Daniel R.  
APPLICANT: Li, Yi  
APPLICANT: Rosen, Craig A.  
APPLICANT: Ruben, Steven M.  
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HPR4J70  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CROCHT,  
ADDRESSEE: STUART & OLSTEIN  
STREET: 6 Becker Farm Road  
CITY: Roseland  
STATE: New Jersey  
COUNTRY: USA

ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/07093  
FILING DATE:  
CLASSIFICATION:

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/465,980  
FILING DATE: 06-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferraro, Gregory D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-446  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1744  
TELEFAX: 201-994-1700  
INFORMATION FOR SEQ ID NO: 1:

## SEQUENCE CHARACTERISTICS:

LENGTH: 1474 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA

## FEATURE:

NAME/KEY: CDS  
LOCATION: 274..1233  
PCT-US95-07093-1

Query Match 12.6%; Score 118.4; DB 5; Length 1474;  
Best Local Similarity 48.9%; Pred. No. 2.6e-28;  
Matches 317; Conservative 0; Mismatches 331; Indels 0; Gaps 0;

QY 109 gtgacacatttggaatcgtgacattatctagtcgaagcctgagacacaaactcat 168  
DB 379 gtggcaatgtgtgaaactgcatgctgttcattcgttaagagacgaacgctgcac 438  
QY 169 accccatgattttttcttaccatctactcctgagcttctgtttacacaaatgt 228  
DB 439 gcttcgaatgaccttcttctgcatgcttgacacgaattgacgcttattccacatcc 498





```

: SOFTWARE: Patent Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/599,252
: FILING DATE: 09-FEB-1996
:
: CLASSIFICATION: 435
:
: ATTORNEY/AGENT INFORMATION:
: NAME: MURASHIGE, KATE H.
: REGISTRATION NUMBER: 29,959
: REFERENCE/DOCKET NUMBER: 9053-0001.21
:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 887-1500
: TELEFAX: (202) 887-0763
:
: TELEX: 90-4030
:
: INFORMATION FOR SEQ ID NO: 84:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 1320 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: US-08-599-252-84

```

Query Match	9.6%	Score 90.2;	DB 1;	Length 1320;
Best Local Similarity	65.8%	Pred. No. 2.9e-19;		
Matches 131; Conservative	0;	Mismatches 66;	Indels 0;	Gaps 0

Accession	Sequence	Position
QY	778 ctgcacacaccttcgcccagctcccaagaagcaagaagaatggtctctctctctatgta	837
Db	1259 CTCGAGCAGCAATAATCTTATTTCCAGGAGCAGGAGGCAATGTTCTTAACCTTTTACACA	1200
QY	838 atcatctgcaccatgctcgtgaatcccttatataacacttaagaaacaaggaggtlaaagaa	897
Db	1199 ATTCTCACTCCGAGTGTTAAACCCCTGATCTATACACTAAGAACAAMAGATCTTAAAGAG	1140
QY	898 ggccttaaaagttggttg 916	
Db	1139 GCCATGGAAGAAGTGCTAG 1121	

RESULT 15  
 US-08-436-074-57/c  
 Sequence 57, Application US/08436074  
 Patent No. 5753438  
 GENERAL INFORMATION:  
 APPLICANT: DRAVNA, DENNIS T.  
 APPLICANT: FEDER, JOHN N.  
 APPLICANT: GINTRE, ANDREAS  
 APPLICANT: KIMMEL, BRUCE E.  
 APPLICANT: THOMAS, WINSTON J.  
 APPLICANT: MOLEF, ROGER K.  
 TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY  
 TITLE OF INVENTION: HEROCROMATOSIS  
 NUMBER OF SEQUENCES: 57  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: MORRISON & FOERSTER  
 STREET: 2000 pennsylvania Ave. N.W., Suite 5500  
 CITY: Washington  
 STATE: DC  
 COUNTRY: USA  
 ZIP: 20006-1888  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/436,074  
 FILING DATE: 08-MAY-1995  
 CLASSIFICATION: 436

ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 9053-0001.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ. ID NO.: 57:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1320 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-436-074-57

Query Match	9.6%;	Score 90.2;	DB 1;	Length 1320;
Best Local Similarity	65.8%;	Pred. No. 2,9e-19;		
Matches 131; Conservative	0;	Mismatches 66;	Indels 0;	Gaps 0;

QY	778	ctcgaaacccacttcgcccagctcccaagaaccaaagaagaatggtctctctctctatgta	837
Db	1259	cttcagcagccacgaactactttttcccccagagacgaaggcaattttctttacacaa	1200
QY	838	atcattgcacccatgctcgtgaatcccttatataacacttaggaacacaaggaggttaaggaa	897
Db	1199	attttctactctccacgtgttaaacccccctgacttatacacttagaanaacaaagatgtttaaag	1140
QY	898	ggctttaaagggttggtg	916
Db	1139	gccattgacgaaggctgctag	1121

```
Search completed: February 26, 2002, 05:02:56
Job time: 4895 sec
```





```

/gene="dJ408B20.1"
/note="dJ408B20.1 (novel 7 transmembrane receptor
(factory family) (hs6m1-33p) pseudogene)
match: CDNA: Em:M64382 Em:AF102516 Em:Y15524 Em:Y15525
Em:X89699 Em:AF102521 Em:AF042016 Em:AF044033 Em:AF102525
Em:AF044034 Em:AF044035 Em:AJ003145 Em:Y14442 Em:AF044037
Em:Y10529 Em:AF102529 Em:AF044038 Em:AF091565 Em:Y07557
Em:X89706 Em:AF042020 Em:M64391 Em:AF102532 Em:AF102533
Em:M64376 Em:AF102535 Em:M64377 Em:M64378 Em:AF102537
Em:AF091573 Em:AF091574 Em:AF091577 Em:AF091578
Em:AF102540 Em:M64385 Em:AF044053 Em:M64386 Em:M64387
Em:M64388 Em:AF091580 Em:DI2820
match: proteins: Tr:076000 Tr:076001 Tr:076002 Tr:035434
Sw:Q15062 Tr:Q95918 Tr:Q9TUB8 Tr:Q9S006 Tr:Q9Y339
Tr:Q9WV13 Tr:Q95918 Tr:Q9TUB8 Tr:Q9S006 Tr:Q9Y339
Tr:Q9Q222 Sw:Q95156 Sw:Q95157 Tr:Q9UGF6 Tr:Q9Y239
Tr:Q9UJ15 Tr:Q62007 Sw:P23275 Tr:Q95371 Tr:Q9UJ20
Tr:Q9Q218 Tr:Q9Q219 Tr:Q9PSJ3 Tr:P70526 Tr:Q9Q221"
/codon_start=1
/pseudo
evidence=not_experimental
complement(1518, 1963)
/note="match: GSS: Em:AQ456128"
2024, 2298
repeat_region
/note="Aluud repeat: matches 1, .280 of consensus"
2299, 2382
repeat_region
/note="42 copies 2 mer ag 79% conserved"
2559, 2650
repeat_region
/note="2 copies 46 mer 91% conserved"
complement(4227, 4576)
/misc_feature
/note="match: STS: Em:G31214"
4719, 5074
repeat_region
/note="L2 repeat: matches 2375, .2750 of consensus"
complement(4834, 5586)
/misc_feature
/note="match: GSS: Em:AQ748194"
5098, 5236
repeat_region
/note="L2 repeat: matches 1381, .1525 of consensus"
5707, 5843
repeat_region
/note="MIR repeat: matches 102, .240 of consensus"
6275, 6509
repeat_region
/note="L1M4 repeat: matches 5093, .5329 of consensus"
complement(6590, 6889)
/misc_feature
/note="match: GSS: Em:U94441"
6679, 6752
repeat_region
/note="MER76 repeat: matches 613, .685 of consensus"
6753, 6850
repeat_region
/note="L1P repeat: matches 1, .98 of consensus"
6820, 7010
repeat_region
/note="L1P repeat: matches 580, .776 of consensus"
7006, 7615
repeat_region
/note="L1P repeat: matches 900, .1509 of consensus"
7611, 9145
repeat_region
/note="L1P repeat: matches 3813, .5344 of consensus"
9751, 10411
repeat_region
/note="HERVL repeat: matches 3244, .3919 of consensus"
10555, 10663
repeat_region
/note="L1R16A repeat: matches 203, .307 of consensus"
10964, 11131
repeat_region
/note="L1R1H repeat: matches 1, .168 of consensus"
11705, 12141
repeat_region
/note="L1M4 repeat: matches 3898, .4367 of consensus"
13143, 13516
/misc_feature
/note="match: GSS: Em:AQ100659"
13668, 13803
repeat_region
/note="MIR repeat: matches 58, .199 of consensus"
complement(15225, 15506)
/misc_feature
/note="match: GSS: Em:AQ100750"
15627, 16228
misc_feature
/note="match: GSS: Em:AQ544678"
16824, 17089
repeat_region
/note="Aluub repeat: matches 29, .311 of consensus"
17090, 17158
repeat_region
/note="Aluuf/FLAM repeat: matches 1, .81 of consensus"

repeat_region 17553, 17848
/note="AluX repeat: matches 3, .298 of consensus"
17789, 18240
misc_feature
/note="match: STS: Em:G31213"
20021, 20249
repeat_region
/note="L1P1 repeat: matches 5904, .6155 of consensus"
20708, 21649
gene
20708, 21649
/gene="dJ408B20.2"
CDS
20708, 21649
/note="match: CDNA: Em:AF044034 Em:Y14442 Em:X89700
Em:Y07557 Em:AF042023 Em:AF091573 Em:AF091574 Em:X89670
Em:M64391 Em:AF102540 Em:AF010293 Em:M64388 Em:DI2820
Em:AF034896
match: ESTs: Em:N68399
match: proteins: Tr:Q43883 Tr:Q9Q221 Tr:Q9Z1K0 Tr:Q9Q220
Tr:Q63384 Tr:Q95439 Tr:Q9Q219 Sw:P23270 Tr:Q9Y387
Tr:Q9Q222 Tr:Q9Y299 Tr:Q62943 Tr:Q90807 Tr:Q90808
Tr:Q9WV13 Tr:Q63395 Tr:Q95918 Sw:P23266 Sw:P37069
Sw:Q95156 Tr:Q9UGF4 Tr:Q9WU86 Tr:Q60403 Tr:Q76000
Sw:P37070 Tr:Q9QW38 Tr:Q77758 Tr:Q9Q217 Tr:Q9R0K3
Tr:Q9R0K4 Tr:Q60412 Tr:Q9QY00 Tr:Q9QW36 Tr:Q9UJ20
Sw:P37072 Sw:P37071 Tr:Q9Y385"
/codon_start=1
evidence=not_experimental
/product="dJ408B20.2 (novel 7 transmembrane receptor
(factory family) (hs6m1-33))"
/misc_feature
/protein_id="CAC14158.1"
/db_xref="GI:10944516"
/translaton="MNWVDSIIQIEFILLGFSRPMLEPPLLVFLISYTVTIFGNLT
ILVSRIDTQLHTPMYEFPLTINSLIDLTCTCTPOMLVNLCISIRKYISYGCVAOLF
IFLALGATEYLLAVMSFDRVAICRPVAVIMHORCLQLAASWVTGFSNVMLS
TUTLQPLCDPYVIDHFEICFEPALCKISVETNEALFVSELPFILITLILISY
AFIVRAVIRIOSAGIDHRCARCTGCSHILVLEFSTAASVVLQPPSPSKQCKMNVSL
FGYIAPMANLITITLRNKEVKEGKRLVAAYFLIKK"
21073, 21162
repeat_region
/note="L1R15 repeat: matches 127, .212 of consensus"
complement(21244, 21671)
/misc_feature
/note="match: GSS: Em:AQ077154"
21964, 21999
repeat_region
/note="18 copies 2 mer ca 100% conserved"
23289, 23944
repeat_region
/note="L1M2 repeat: matches 5508, .6168 of consensus"
24195, 24291
repeat_region
/note="FLAM C repeat: matches 38, .136 of consensus"
24448, 24781
/misc_feature
/note="match: GSS: Em:B75689"
24537, 24857
repeat_region
/note="match: GSS: Em:AQ102724"
24739, 25051
repeat_region
/note="Alusg repeat: matches 1, .310 of consensus"
25407, 25944
repeat_region
/note="L1M4 repeat: matches 5206, .5763 of consensus"
25945, 26050
repeat_region
/note="Alusg/x repeat: matches 195, .302 of consensus"
26055, 26632
misc_feature
/note="match: GSS: Em:AQ057601"
26069, 26126
repeat_region
/note="L1M57 repeat: matches 6111, .6169 of consensus"
26164, 26220
repeat_region
/note="L1M57 repeat: matches 6111, .6168 of consensus"
26221, 26391
repeat_region
/note="Aluub repeat: matches 137, .306 of consensus"

alignment_scores:
Quality: 1601.00 Length: 313
Ratio: 5.115 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-755-017-2 x AL133267
Align seg 1/1 to: AL133267 from: 1 to: 44788

```

1 MetAsnTrpValAlaAspSerIleIleGlnGluPheIleLeuGly 17  
 20708 ATGATTTGGGTAATGACAGCATCATACAGGATTTATTCCTGGGTTT 20757  
 17 eSerAspArgProIleGluPheProLeuValValPheLeuIle 34  
 20758 CTCAGATCGACCTTGGCTGGAGTTCCACTCTTGGCTCTTGATTT 20807  
 34 eTyrThrValIleThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50  
 20808 CTTACACGTGACCACTTGGCAATCGACCATTAATTCATGTCACAGC 20857  
 51 LeuAspThrIleLeuHisThrProMetTyrPhePheLeuThrAsnLeu 67  
 20858 CTGGACACCAACTTCATACCCCGCATGATTTTCTTACCAATCTATC 20907  
 67 rLeuLeuAspLeuGlyCysTyrThrThrCysThrValProGlnMetLeuVal 84  
 20908 ACTCGTGATCTTTGTACACCATGTACAGTCCACCAATGCTAGTAA 20957  
 84 snLeuGlySerIleArgIleValIleSerTyrArgGlyCysValAlaGln 100  
 20958 ATTATGACAGCATCAGAAAGTAATCAGTATCGCTGTGTACCCAG 21007  
 101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuAlaVal 117  
 21008 CTTTTCATATTTCTGGCTTGGGGCTACTGATATCTTCTCTGGCGCT 21057  
 117 MetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSerV 134  
 21058 CATGCCCTTTGATAGTTGTAGCATTTTGGGCTCTCCATTTACTCAG 21107  
 134 aIleMetHisGlnArgLeuGlyLeuGlnLeuAlaAlaSerTrpVal 150  
 21108 TTATCATCTACACAGACTGCTGCCAGTGGCAGCTGCATCTCGGGT 21157  
 151 ThrGlyPheSerAsnSerValTyrPheSerThrIleThrLeuGlnLeu 167  
 21158 ACTGTTTGTAGTACTGAGTGGTGTCTCCCTGACTCTCCAGCTGCC 21207  
 167 OLeuGlyAspProTyrValIleAspHisPheLeuGlyGluValProAla 184  
 21208 ACTCGTGACCCCTATGTGATAGATCATCTTCTCTGTAAGCTCCAGC 21257  
 184 eLeuLeuGlyLeuSerCysValGluThrThrAlaAsnGlnAlaGluLeu 200  
 21258 TGCTCAAGTTATCTTGTGTGACACACAGCAATAGAGGCTGAACTATTC 21307  
 201 LeuValSerGluLeuPheHisLeuIleProLeuThrLeuIleLeuIle 217  
 21308 CTTGTCAGTGAAGCTCTCCATCTAATACCCCGACACTCATCTTATATC 21357  
 217 rTyrAlaPheIleValAlaArgAlaValIleuArgIleGlnSerAlaGlu 234  
 21358 ATATGCTTTTATNTGTCCGACGAGTATGAGGATACAGTCTGTAAGTTC 21407  
 234 rGlnIleValAlaPheGlyThrCysGlySerHisLeuIleValValSerLeu 250  
 21408 GACAAAAAGCAATTTGGACATCTGCTCCATCTAATGTTGCTGCTCTT 21457  
 251 PheTyrSerThrAlaValSerValTyrLeuGlnProProSerProSerSe 267  
 21458 TTTTATATGACGCGCTCTGTGTACCTGCAACCACTTCCGCCAGCTC 21507  
 267 rLysAspGlnGlyLysMetValSerLeuPheTyrGlyIleIleAlaPro 284  
 21508 CAAGGACCAAGAAAGATGTTTCTCTCTATGGAATCATTTGACACCA 21557  
 284 eLeuAsnProLeuIleTyrThrLeuArgAsnGlyGluValLysGluGly 300  
 21558 TGCTGAATCCCTTATATATACACTTAGACACCAAGAGGTAAGGAAGC 21607

301 PheLysArgLeuValAlaArgValPheLeuIleLysLys 313  
 21608 TTTAAAGCTTGCTGCAAGAGCTCTTCTTATACAGAAA 21646

seq\_name: gb\_hlg:AC025941

seq\_documentation\_block: 176277 bp DNA HNG 26-MAY-2000  
 LOCUS AC025941  
 DEFINITION Homo sapiens chromosome 6 clone RP11-635011 map 6, WORKING DRAFT  
 SEQUENCE, 15 unordered pieces.  
 AC025941  
 AC025941.2 GI:7408057  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 1 (bases 1 to 176277)  
 Birren, B., Linton, L., Nussbaum, C., Lander, E., Abrahams, H., Allen, N.,  
 Anderson, S., Baldwin, J., Baran, N., Bastien, V., Bedalov, F.,  
 Boguslavsky, L., Bouckhelter, B., Brown, A., Burkett, G.,  
 Campiano, A., Castle, A., Chepel, Y., Colangelo, M., Collins, S.,  
 Collymore, A., Cooke, P., DeArillano, K., Dewar, K., Diaz, J.S.,  
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,  
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
 Grand-Pierre, N., Grant, G., Hages, B., Heatford, A., Horton, L.,  
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
 Klein, J., Lacombe, K., Lamazares, R., Landers, T., Lebeck, J.,  
 Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,  
 McCarthy, M., McEwan, P., McGuck, A., McKernan, K., McPeckers, R.,  
 Melchior, J., Meneses, L., Milroy, T., Miranda, C., Minge, V., Morrow, J.,  
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
 O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rotman, D.,  
 Roy, A., Santos, R., Schauer, S., Severly, P., Spencer, B.,  
 Stange-Rhoman, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Teste, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,  
 Vassiliev, H., Veli, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

JOURNAL  
 COMMENT  
 Submitted (17-MAR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 4, 2000 this sequence version replaced gi:7259782.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu

Project Information

Center project name: L8577  
 Center code: 635-O-11  
 Center clone name: 635-O-11  
 Summary Statistics  
 Sequencing vector: M13; M7815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 168925 bases at least Q40  
 Consensus quality: 172827 bases at least Q30  
 Consensus quality: 174151 bases at least Q20  
 Insert size: 176000; agarose-efp  
 Insert size: 174877; sum-of-ctrls  
 Quality coverage: 4.5 in Q20 bases; agarose-efp  
 Quality coverage: 4.6 in Q20 bases; sum-of-ctrls

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 15 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 2997: contig of 2997 bp in length.  
\* 2998 3097: gap of 100 bp  
\* 3098 5606: contig of 2509 bp in length  
\* 5607 5706: gap of 100 bp  
\* 5707 9209: contig of 3503 bp in length  
\* 9210 9309: gap of 100 bp  
\* 9310 13614: contig of 4305 bp in length  
\* 13615 13714: gap of 100 bp  
\* 13715 20639: contig of 6925 bp in length  
\* 20640 20739: gap of 100 bp  
\* 20740 26545: contig of 5806 bp in length  
\* 26546 26645: gap of 100 bp  
\* 26646 37068: contig of 10423 bp in length  
\* 37069 37168: gap of 100 bp  
\* 37169 48083: contig of 10915 bp in length  
\* 48084 48183: gap of 100 bp  
\* 48184 59015: contig of 10832 bp in length  
\* 59016 59115: gap of 100 bp  
\* 59116 70420: contig of 11305 bp in length  
\* 70421 70520: gap of 100 bp  
\* 70521 80466: contig of 9946 bp in length  
\* 80467 80566: gap of 100 bp  
\* 80567 92325: contig of 11759 bp in length  
\* 92326 92425: gap of 100 bp  
\* 92426 110572: contig of 18147 bp in length  
\* 110573 131898: contig of 21226 bp in length  
\* 131899 131998: gap of 100 bp  
\* 131999 176277: contig of 44279 bp in length.

## FEATURES

source  
1. 176277  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6"  
/clone="RP11-635011"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. 2997  
/note="assembly\_fragment"  
3098. 5606  
/note="assembly\_fragment"  
clone\_end:sp6  
vector\_side:left"  
5707. 9209  
/note="assembly\_fragment"  
9310. 13614  
/note="assembly\_fragment"  
13715. 20639  
/note="assembly\_fragment"  
20740. 26545  
/note="assembly\_fragment"  
clone\_end:r7  
vector\_side:right"  
26646. 37068  
/note="assembly\_fragment"  
37169. 48083  
/note="assembly\_fragment"  
48184. 59015  
/note="assembly\_fragment"  
59116. 70420  
/note="assembly\_fragment"  
70521. 80466  
/note="assembly\_fragment"  
80567. 92325  
/note="assembly\_fragment"  
92426. 110572  
/note="assembly\_fragment"  
110673. 131898

/note="assembly\_fragment"  
131999. 176277  
misc\_feature /note="assembly\_fragment"  
BASE COUNT 54493 a 33357 c 33254 g 53772 t 1401 others  
ORIGIN

alignment\_scores:  
Quality: 1601.00 Length: 313  
Ratio: 5.115 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-755-017-2 x AC025941/rev

Align seg 1/1 to reverse of: AC025941 from: 1 to: 176277

1 MetAsnTrpValAsnAspSerIleIleGlnGluPheIleLeuLeuGlyPhe 17

154174 ATGAATTGGGTAATGACAGCATCATACAGAGATTATTCTGCTGGTTT 154125

17 eSerAspArgProTrpLeuGluPheProLeuValAlaPheLeuIleS 34

154124 CTCAGATCGACCTTGGCTGAGTTTCCACTCTGTGGCTCTTGATT 154075

34 eTrpThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50

154074 CTACACGTGACCATCTTGGCAATCTGACCATTTATCTAGTCGACCC 154025

51 LeuAspThrIleLeuHisThrProMetIyrPhePheLeuThrAsnLeuS 67

154024 CTGGACACCAAACTTCATACCCCATGTATTTTCTTACCAATCTATC 153975

67 rIleuLeuAspIleuCysTrpThrThrcysThrValProGlnMetLeuVal 84

153974 ACTCCTGGATCTTTGTTACACCATGATACAGTCCACAAATGCTAGTA 153925

84 snLeuCysSerIleArgIleValIleSerIyrArgIleCysValAlaGln 100

153924 ATTATGACAGCATCAGCAAGATATCATGTTATCGTGGCTGATGCCCG 153875

101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTrpLeuLeuAlaVal 117

153874 CTTTTCATATTTCTGGCTTGGGGCTGACTGAAATATCTCTCGTGGCC 153825

117 lMetSerPheAspArgPheValAlaIleCysArgProLeuHisIyrSerV 134

153824 CATGCTCTTGTAGAGGTTGTAGCTATTTGTGGCTCTCCATTACTCAG 153775

134 alIleMetHisGlnArgLeuCysLeuGlnLeuAlaAlaIleSerTrpVal 150

153774 TTATCATGACACGACGACACTGCTGCCTGACCTGACACTGCTGGTT 153725

151 ThrGlyPheSerAsnSerValTrpLeuSerThrLeuThrLeuGlnLeuP 167

153724 ACTGGTTTATGATACCTGAGTGTGTTCTACCTGACACTCCAGCTGCC 153675

167 oleuCysAspProTrpValIleAspHisPheLeuCysGluValProAla 184

153674 ACTGTGTGACCCCTATGTGATGATCACTTCTGTGAGTCCCTGGCAC 153625

184 euleuLeuSerCysValGluThrThrAlaAsnGluAlaGluLeuPhe 200

153624 TCCTCAAGTTATCTGTGTGAGACACAGCAAAATGAGGCTCAACTATTC 153575

201 LeuValSerGluLeuPheHisLeuIleProLeuThrLeuIleLeuIleS 217

153574 CTTGTCAAGTACCTCTCCATCTAATACCCCTGACACATCATCTATATC 153525

217 rTrpAlaPheIleValArgAlaValLeuArgIleGlnSerAlaGluGlyA 234

153524 ATATGCTTTTATTTGTCGAGCAGTATGAGATACAGTCTGCTGAAGTTC 153475

234 rglInuysAlaphThCysGlySerHisLeuValSerLeu 250  
|||||  
153474 GACAAAACATTGGGACATGGTCCCATCTATTTGGTCTCTT 153425  
251 pheTySerThAlaValSerValTyLeuGlnProPseProSe 267  
|||||  
153424 TTTTATAGACAGCCGCTCTGTGACCAACCACTCCGCGAGCTC 153375  
267 rlySAspGlnGlyLysMetValSerLeuPheTyGlylleAlaP 284  
|||||  
153374 CAAGGCCAGGAAAGATGGTTCTCTCTTAATGGAATCATTCACCA 153325  
284 etLeuAsnProLeuLeuTyThrLeuArgAsnLysGluValLysG 300  
|||||  
153324 TCGTGATCCCTTATATATACCTTAGCAACAGAGTAAGAGAGC 153275  
301 pheLysArgLeuValAlaArgValPheLeuLeuLysLys 313  
|||||  
153274 TTTAAAGGTGGTGCAGAGCTCTTAATCAAGAA 153236  
seq\_name: gb\_pr: HSA302593  
seq\_documentation\_block:  
LOCUS HSA302593 1074 bp DNA PRI 06-JAN-2001  
DEFINITION Homo sapiens 6M1-10\*02 gene for olfactory receptor, cell 11ne  
KR3598.  
ACCESSION AJ302593  
VERSION AJ302593.1 GI:12054410  
KEYWORDS 6M1-10\*02 gene; olfactory receptor.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1074)  
AUTHORS Ziegler, A., Ehlers, A., Forbes, S. A., Trowsdale, J.,  
Uchanska-Ziegler, B., Volz, A., Younger, R., and Beck, S.  
TITLE Polymorphic olfactory receptor genes and HLA loci constitute  
extended haplotypes  
(in) Kasahara, M. (Ed.):  
MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND  
FUNCTION: 110-130;  
Springer-Verlag, Tokyo, Japan (2000)  
REFERENCE 2 (bases 1 to 1074)  
AUTHORS Ziegler, A., Ehlers, A., Forbes, S. A., Trowsdale, J., Volz, A.,  
Younger, R. M., and Beck, S.  
TITLE Polymorphisms in olfactory receptor genes: a cautionary note  
Unpublished  
3 (bases 1 to 1074)  
AUTHORS Younger, R. M., Amadou, C., Bethel, G., Ehlers, A., Fischer, Lindahl, K.,  
Forbes, S. A., Horton, R., Milne, S., Mungall, A. J., Trowsdale, J.,  
Volz, A., Ziegler, A., and Beck, S.  
TITLE Characterisation of clustered MHC-linked Olfactory Receptor Genes  
in Human and Mouse  
Unpublished  
4 (bases 1 to 1074)  
AUTHORS Ehlers, A., Beck, S., Forbes, S. A., Trowsdale, J., Volz, A., Younger, R.,  
and Ziegler, A.  
TITLE MHC-linked olfactory receptor loci exhibit polymorphism and  
contribute to extended HLA/OR-haplotypes  
Unpublished  
5 (bases 1 to 1074)  
AUTHORS Ehlers, A.  
TITLE Direct Submission  
Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik,  
Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,  
GERMANY  
FEATURES  
Source location/Qualifiers  
1..1074  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/cell\_line="KR3598"  
gene 1..1074

/gene="6M1-10\*02"  
1..1074  
/gene="6M1-10\*02"  
/codon\_start=1  
/product="olfactory receptor"  
/protein\_id="CAC20513.1"  
/db\_xref="GI:12054411"  
/translation="MNNVKNKSVPEFLIVSDPQMLEIPFVWFLESYITITFIGNLT  
ILVSHDFKLIHTPWFPLSLJSLDLCYTSYVPMVLNICTNRYISIGGVAAQLP  
IFLAGSTECILLAVMCFDFEVALICRPLHSI IMHQRICEQLAASMTSGFSNSVLDS  
TWLTKMPLCGHKEVDHFECEVPALCKSCVDPTNNEALEFISVLPIPTLILISY  
AFIVQAVLRIOASGRRKAEFCGSHLIVSLFYGTAISMVLOPPSSKRGKMWSL  
FCGLIAPMLNPLIYTLNRKYEKFAFKRLVANSLSLNOELRNQMISFAKDIYLYLTNPF  
SASCPPIVYTIENCNLPORFEP"  
BASE COUNT 277 a 248 c 203 g 346 t  
ORIGIN  
Alignment\_scores:  
Quality: 1323.00 Length: 314  
Ratio: 4.531 Gaps: 1  
Percent Similarity: 92.994 Percent Identity: 82.803  
alignment\_block:  
US-09-755-017-2 x HSA302593 ..  
Align seg 1/1 to: HSA302593 from: 1 to: 1074  
1 MetAsnTrpValAsnAspSerIleIleGlnGluPheIleLeuGluLys 17  
|||||  
1 ATGAATTTGGGTAATAGACTGTCGCCAGAGATTCATCTCTGACTTTT 50  
17 eSerAspArgProTrpLeuGluPheProLeuValAlaPheLeuLys 34  
|||||  
51 CTCAGATCAACCATGGCTAGAGATTCACACCTTTGATGATTTCTGTTT 100  
34 eTrpThrValThrIlePheGlyAsnLeuThrIleLeuValSerArg 50  
|||||  
101 CCTATATCTTGACAAATCTGGCAATCTGCACATTAATCTTGTCACAT 150  
51 LeuAspThrLysLeuHisThrProMetTyrPhePheLeuThrAsnLeu 67  
|||||  
151 GTGGATTTCACAACTCCACACCCCTATGACTTTTCTTAGACATCTCTC 200  
67 rLeuLeuAspLeuCysTyThrThrCysThrValProGlnMetLeuVal 84  
|||||  
201 ACTCTGAGACCTTTGCTATACCAAGTACAGTCCACAATCTGCTGTA 250  
84 snLeuGlySerIleArgLysValIleSerTyArgGlyCysValAlaGln 100  
|||||  
251 ACATATGCAACACACAGGAAAGTATCATGTTAGTGCTGTGGCCAG 300  
101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyLeuLeuLeuAla 117  
|||||  
301 CTTTATTTCTCTGGCTTTGGCTTGGTTCACAGAAATGCTTCTCTGGCCG 350  
117 lMetSerPheAspArgPheValAlaIleCysArgProLeuHisTySer 134  
|||||  
351 CATGCTCTTGATAGAGTTGTACCTATTGTGCGCCTCCCATTTACTCAA 400  
134 alLeuMetHisGlnArgLeuCysLeuGlnLeuAlaAlaAspTrpVal 150  
|||||  
401 TTATCATGACACGAGGCTGCTTCAGTTGGACCTGCATCTGCATTT 450  
151 ThrGlyPheSerAsnSerValTrpLeuSerThrLeuThrLeuGlnLeu 167  
|||||  
451 AGTGGCTTTAGCATTCAGATTACAGTCCACCTGCAGCACTTAAGATGCC 500  
167 oLeuCysAspProTyTrValIleAspHisPheLeuCysGluValProAla 184  
|||||  
501 ACTGTGTGTGCACAAAGAGTGGATCATCTCTGTGAAGTCCCTGCTC 550  
184 euLeuTyLeuSerCysValGluThrThrAlaAsnGluAlaGluLeuPhe 200

```

|||||
551 TGCCAGAGTGTCTCTGTGACACACACAAATGAGCGTGAACATATTC 600
201 LeuValSerGluLeuPheHisLeuIleProLeuThrLeuLeuIle 217
601 TTTATGCTGTCTATTCCTTCTATACCCGTGACCTCATCTTATATC 650
217 TTTATGCTGTCTATTCCTTCTATACCCGTGACCTCATCTTATATC 234
651 GTATGCTTTATTTGTCACACAGTGTGAGATTCACCTGCTGAGAGTC 700
234 rglLysAlaPheGlyThrCysGlySerHisLeuIleValSerLeu 250
701 GACGAAGCGATTTGGACATGTGGCTCCATCATTAATGTGTGACTT 750
251 PheTyrSerThrAlaValSerValTyrLeuGlnProProSerProSer 267
751 TTTATGCTGTCTATTCCTTCTATACCCGTGACCTCATCTTATATC 800
267 rLysAspGlnGlyLysMetValSerLeuPheTyrGlyIleLeuAlaPhe 284
801 CAAGACCGGGGAAAGATGGTTCTCTCTGTGAGATCATTCGACCCA 850
284 etLeuAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGly 300
851 TGCTGAATCCCTTATATATACACTTAGCAACAAAGAGTAAGAGACC 900
301 PheLysArgLeuValAlaArgValPheLeuIleLys 313
901 TTTAAAGGTTGTTGCCAAGAGCTTCTTATATCAGAAA 940
seq_name: gb_pr: HSA302584
seq_documentation_block: 1074 bp DNA PRI 06-JAN-2001
LOCUS HSA302584
DEFINITION Homo sapiens 6M1-10*01 gene for olfactory receptor, cell line
BM28.7.
ACCESSION AJ302584
VERSION AJ302584.1 GI:12054392
KEYWORDS 6M1-10*01 gene; olfactory receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1074)
Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J.,
Uchanska-Ziegler,B., Volz,A., Younger,R. and Beck,S.
Polymorphic olfactory receptor genes and HLA loci constitute
extended haplotypes
(in) Kasahara,M. (Ed.):
MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND
FUNCTION: 110-130:
Springer-Verlag, Tokyo, Japan (2000)
2 (bases 1 to 1074)
Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J., Volz,A.,
Younger,R.M. and Beck,S.
Polymorphisms in olfactory receptor genes: a cautionary note
Unpublished
3 (bases 1 to 1074)
Younger,R.M., Amdou,C., Bethel,G., Ehlers,A., Fischer,Lindahl,K.,
Forbes,S.A., Horton,R., Milne,S., Mungall,A.J., Trowsdale,J.,
Volz,A., Ziegler,A. and Beck,S.
Characterisation of clustered MHC-linked Olfactory Receptor Genes
in Human and Mouse
Unpublished
4 (bases 1 to 1074)
Ehlers,A., Beck,S., Forbes,S.A., Trowsdale,J., Volz,A., Younger,R.
and Ziegler,A.
MHC-linked olfactory receptor loci exhibit polymorphism and
contribute to extended HLA/OR-haplotypes
Unpublished
5 (bases 1 to 1074)
Ehlers,A.
JOURNAL
REFERENCE
AUTHORS

```

```

TITLE Direct Submission
JOURNAL Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik,
Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,
GERMANY
FEATURES
source Location/Qualifiers
1..1074
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/cell_line="BM28.7"
1..1074
/gene="6M1-10*01"
1..1074
/gene="6M1-10*01"
/codon_start=1
/product="olfactory receptor"
/protein_id="CAC20504.1"
/db_xref="GI:12054393"
/transcript="NM005043"
/translation="MNMVNVKSVPOEFLIVFSDQPMLEIPPFVNFLEFSLITLFGNLT
ILVSHDFKLTHTPMYFELSNLSLDLCYTSVSTVIMORPCPOLAASVTSQSVNLOF
IFALGSTRECLLIAMCFDRFAVTCRPHYSITIMORPCPOLAASVTSQSVNLOF
TWTLKMPICGKREVDHFEVCEPALIKLSVDTTANAEALFISVLFILPTLILIS
AFIVDAVLRIGSABGQRKAFGTCGSHLIVSLFTGTALSMILOPSPSKRGKMWSL
FCGIAPMLNPLIYTLRNKREKAEFRILVARSLLNOETIRNMOMISFAKDTLVLTNPF
SASCPIFVITTEYCNILPORKFP"
BASE COUNT 278 a 248 c 202 g 346 t
ORIGIN
alignment_scores:
Quality: 1319.00 Length: 314
Ratio: 4.517 Gaps: 1
Percent Similarity: 92.994 Percent Identity: 82.484
alignment_block:
US-09-755-017-2 x HSA302584 ..
Align seg 1/1 to: HSA302584 from: 1 to: 1074
1 MetAsnTrpAlaAsnAspSerIleIleGlnGluPheLeuLeuGlyPh 17
|||||
1 ATGAATTGGGTAAATAGAGTGTCCACAGAGACTTATCTGTAGATT 50
17 eSerAspArgProTrpLeuGlnPheProLeuValAlaPheLeu 34
|||||
51 CTGAGATCAACCATGGCTAGAGATTCACCCCTTGATGATGTTCTGTT 100
51 CTGAGATCAACCATGGCTAGAGATTCACCCCTTGATGATGTTCTGTT 100
34 eTyrThrValIlePheGlyAsnLeuThrIleIleLeuValSerArg 50
|||||
101 CTATATCTTGACAACTTGGCAATCTGACAAATTAATCTTGATGATG 150
101 CTATATCTTGACAACTTGGCAATCTGACAAATTAATCTTGATGATG 150
51 LeuAspThrLysLeuHisThrProMetTyrPhePheLeuThrAsnLeu 67
|||||
151 GTGGATTTCAACTCCACACCCCTATGACTTTTCTTAGCAATCTCTC 200
151 GTGGATTTCAACTCCACACCCCTATGACTTTTCTTAGCAATCTCTC 200
67 rLeuLeuAspLeuCysTyrThrThrCysThrValProGlnMetLeuVal 84
|||||
201 ACTCTTGACACTTGGCTATGCCACAGTACAGTTCACAATGCTGTAA 250
201 ACTCTTGACACTTGGCTATGCCACAGTACAGTTCACAATGCTGTAA 250
84 snLeuCysSerIleArgLysValIleSerTyrArgGlyCysValAlaGln 100
|||||
251 ACATATGCAACACACAGGAAGTAATGATGATGCTGTGCGCCAG 300
251 ACATATGCAACACACAGGAAGTAATGATGATGCTGTGCGCCAG 300
101 LeuPheIlePheLeuAlaLeuGlyAlaThrGlyTyrLeuLeuLeuAla 117
|||||
301 CTTTTCATTTTCTGTGGCTTGGGTGCTCCACGAATGTCTTCTGCGCC 350
301 CTTTTCATTTTCTGTGGCTTGGGTGCTCCACGAATGTCTTCTGCGCC 350
117 lMetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSer 134
|||||
351 CATGTGCTTGATAGGTTTACATATTTGTGCGCTTCATTTACTCAA 400
351 CATGTGCTTGATAGGTTTACATATTTGTGCGCTTCATTTACTCAA 400
134 alIleMetHisGlnArgLeuCysLeuGlnLeuAlaAlaSerTrpVal 150
|||||

```



401 TTATCATGCCACGAGCTGCTGCTTCACAGTGCACCTCATCTGATT 450  
 151 ThrGlyPheSerAsnSerValTrpLeuSerThrLeuThrLeuGlnLeu 167  
 451 AGTGGCTTTAGCAATTTCAGTATTACAGTCACCTGCACCTTAAGATGCC 500  
 167 oLeuGlyAspProTyrValIleAspHisPheLeuGlyValProAla 184  
 501 ACTGTGTGTCACAAAGAGTGCATCTTCTGTGTGTAAGTCCCTGCTC 550  
 184 euleuGlyLeuSerCysValGluThrThrAlaAsnGluAlaGluLeu 200  
 551 TGCTCAAGTGTCTGTGTGTGACACAGCAATGAGGCTGACATATTC 600  
 201 LeuValSerGluLeuPheHisLeuIleProLeuThrLeuIleLeu 217  
 601 TTCATCAGTGTCTATTCCTTAATACCCGACACCTCATCTATATAC 650  
 217 TyrAlaPheIleValArgAlaValLeuArgIleGlnSerAlaGluGly 234  
 651 GATATGCTTTATGTCACAGCAGTGTGAGAAATCCAGTCTGCTGAGGTC 700  
 234 rglGlnAlaPheIleGlyThrCysGlySerHisLeuIleValSerLeu 250  
 701 AACGAAAGCATTTGGACATGTGGCTCCCATCTATTTGTGTGTCATTT 750  
 251 PheTyrSerThrAlaValSerValTyrLeuGlnProProSerProSer 267  
 751 TTTTATGAGCAGCTATCTCATGACGCAACCCCTTCCACCAAGCTC 800  
 267 rTysAspGlnGlyLeuMetValSerLeuPheTyrGlyIleIleAlaPro 284  
 801 CAAAGACCGGGAAGATGTTCTCTCTGTGTGAAATCATTTGCACCA 850  
 284 eLeuAsnProLeuIleTyrThrLeuArgAsnGlyValGlyGly 300  
 851 TGTGATATCCCTTATATATACCTTAGAACAAGAGTAAGAAGGCC 900  
 301 PheLysArgLeuValAlaArgValPheLeuIleLysLys 313  
 901 TTTTAAAGGTGGTGCACAAAGAGTCTTCTTAATCAAGAA 940

seq\_name: gb\_HSA302585  
 seq\_documentation\_block:  
 LOCUS HSA302585 1074 bp DNA PRI 06-JAN-2001  
 DEFINITION Homo sapiens 6M1-10\*01 gene for olfactory receptor, cell line  
 BM19.7.  
 ACCESSION AJ302585  
 VERSION AJ302585.1 GI:12054394  
 KEYWORDS 6M1-10\*01 gene; olfactory receptor.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 1074)  
 Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J.,  
 Uchanska-Ziegler,B., Volz,A., Younger,R. and Beck,S.  
 Polymorphic olfactory receptor genes and HLA loci constitute  
 extended haplotypes  
 (in) Kasahara,M. (Ed.):  
 MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND  
 FUNCTION: 110-130;  
 Springer-Verlag, Tokyo, Japan (2000)  
 2 (bases 1 to 1074)  
 Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J., Volz,A.,  
 Younger,R.M. and Beck,S.  
 Polymorphisms in olfactory receptor genes: a cautionary note  
 Unpublished  
 3 (bases 1 to 1074)  
 Younger,R.M., Anadou,C., Bethel,G., Ehlers,A., Fischer,Lindahl,K.,  
 Forbes,S.A., Horton,R., Milne,S., Mungall,A.J., Trowsdale,J.,  
 Volz,A., Ziegler,A. and Beck,S.

TITLE Characterisation of clustered MHC-linked Olfactory Receptor Genes  
 in Human and Mouse  
 JOURNAL Unpublished  
 REFERENCE 4 (bases 1 to 1074)  
 AUTHORS Ehlers,A., Beck,S., Forbes,S.A., Trowsdale,J., Volz,A., Younger,R.  
 and Ziegler,A.  
 TITLE MHC-linked olfactory receptor loci exhibit polymorphism and  
 contribute to extended HLA/OR-haplotypes  
 JOURNAL Unpublished  
 REFERENCE 5 (bases 1 to 1074)  
 AUTHORS Ehlers,A.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik,  
 Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,  
 GERMANY

FEATURES  
 source  
 Location/Qualifiers  
 1..1074  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /cell\_line="BM19.7"  
 1..1074  
 /gene="6M1-10\*01"  
 1..1074  
 /gene="6M1-10\*01"  
 /codon\_start=1  
 /product="Olfactory receptor"  
 /protein\_id="CAC20505.1"  
 /db\_xref="GI:12054395"  
 /translation="MMWVNSVPOEFLILVSDPMLPEPVMFLRSYLLTIRGNLT  
 ILVSHVDFKLPMPFLSNLSLIDCYTTSYPMQLVNCNRKYSYGGCYAOLF  
 FLIALSGTECLLAVMFCDFRPAVCRLHLSITHOGLCLOLAASWISPSNVLOS  
 TWTLKMPLOCGRKVDHFCERPAALIKLSCTDTANELEFISVLLIPVTLILSY  
 AFIVQAVLRIOASAGQRARFGTGSGLIVSLFYGTAISWYLOPSPSSKORGMVSL  
 FCGIAPMLNPLIYLRNKEVKAERKRLVAKSLNDEIRNMQMSFAKDVLYLTNFS  
 SASCPDIVITLENCNPORKEP"  
 BASE COUNT 278 a 248 c 202 g 346 t  
 ORIGIN

alignment\_scores:  
 Quality: 1319.00 Length: 314  
 Ratio: 4.517 Gaps: 1  
 Percent Similarity: 92.994 Percent Identity: 82.484

alignment\_block:  
 :US-09-755-017-2 x HSA302585 ..  
 Align seg 1/1 to: HSA302585 from: 1 to: 1074

1 MetAsnTrpValAsnAspSerIleIleGlnGluPheIleLeuGlyPh 17  
 1 ATGAATTTGGTAAATTAAGAGTGTCCACAGAGATTCATCTGTAGTTT 50  
 17 eSerAspArgProTrpLeuGluPheProLeuLeuValAlaPheLeuLys 34  
 51 CTCAGATCAACCACTGGCTAGAGATTCCACCTTTGGAGTGTCTGTTT 100  
 34 eTyrThrValThrIlePheGlyAsnLeuThrIleLeuValSerArg 50  
 101 CCTATATCTTGACAAATCTTGGCAATCTGACATAATTTCTGTGTACAT 150  
 51 LeuAspThrLysLeuHisThrProMetTyrPhePheLeuThrAsnLeu 67  
 151 GTGATTTCAACTCCACACCCCTATGACTTTTCTTAGCAATCTCTC 200  
 67 rLeuLeuAspLeuCysTyrThrThrCysThrValProGlnMetLeuVal 84  
 201 ACTCTGTGACCTTGTCTATACCAAGTACAGTCCACAAATGCTGTAA 250  
 84 snLeuGlySerIleArgLysValIleSerTyrArgLysValAlaGln 100  
 251 ACATATGCAACACGAGAAAGTATCATGTTATGTGTGTGTGTGTGTG 300

101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluThrLeuLeuAla 117  
 301 CTTTTCATTTTCTGGGCTTGGGCTTCACAGAAATGCTTCTCTGGGCGGT 350  
 117 lmetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSer 134  
 351 CATGTGCTGGATAGGTTTGTACTATTGTCGGCCCTCCATTAACA 400  
 134 alIemethSgIArgLeuCysLeuGlnLeuAlaAlaIleSerTyrVal 150  
 401 TTATTCATGCCACAGAGGCTCTGCTCCAGTTGGACGCTGCATCTGGATT 450  
 151 ThrGlyPheSerAsnSerValIrrPleSerThrLeuThrLeuGlnLeuP 167  
 451 AGTGGCTTATGACATTCAGATTACAGTCCACCGTGGACACTTAAGATGCC 500  
 167 oLeuCysAspProTyrValIleAspHisPheLeuCysGluValProAla 184  
 501 ACTGTGTGTGCACAAAGAGTGGATCACTTCTCTGTGAAGTCCCTGCTC 550  
 184 euleuLysLeuSerCysValGluThrThrAlaAsnGluAlaGluLeuPhe 200  
 551 TGTCTCAAGTTGCTCTCTGTGTGACACACACCAATGAGGCTGAACATATTC 600  
 201 LeuValSerGluLeuPheHisLeuIleProLeuThrLeuLeuLeuIle 217  
 601 TTCATCAGTGTGCTATTCCTTAATACCGGTGACACTCATCTTAATATC 650  
 217 rTyrrAlaPheIleValArgAlaValLeuArgIleGlnSerAlaGluGly 234  
 651 GTATGCTTTTATTTGTCACACAGCTGTGAGAAATCCAGTGTGAGAGTCT 700  
 234 rGcIlnLysAlaPheGlyThrCysGlySerHisLeuIleValIleSerLeu 250  
 701 AACGAAAGGATTTGGGACATGTGCTCCCATCTAATGTGGTGCTACTT 750  
 251 PheTyrSerThrAlaValSerValIrrLeuGlnProProSerProSer 267  
 751 TTTTATGATGACGCTATCTCATGATGCTGACACACACTTCCACCAAGCTC 800  
 267 rLysAspGlnGlyLysMetValSerLeuPheTyrGlyIleIleAlaPro 284  
 801 CAAAGCCGGGGAAGATGGTTCCTCTCTGTGGAATCATATTGCAACCA 850  
 284 etLeuAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGluGly 300  
 851 TGTGTGATCCCTTATATATACACTTACGACAAAGAGTAAAGGAGAGCC 900  
 301 PheLysArgLeuValAlaArgValPheLeuIleLysLys 313  
 901 TTTAAAGGTTGGTGCMAAGAGTCTTCTTAATCAAGAAA 940

seq\_name: gb\_pr: HSA302586

seq\_documentation\_block:  
 LOCUS HSA302586 1074 bp DNA PRI 06-JAN-2001  
 DEFINITION Homo sapiens 6M1-10\*01 gene for olfactory receptor, cell line LG2.  
 ACCESSION AJ302586  
 VERSION AJ302586.1 GI:12054396  
 KEYWORDS 6M1-10\*01 gene; olfactory receptor.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1074)  
 AUTHORS Ziegler, A., Ehlers, A., Forbes, S.A., Trowsdale, J.,  
 Uchanska-Ziegler, B., Volz, A., Younger, R., and Beck, S.  
 TITLE Polymorphic olfactory receptor genes and HLA loci constitute  
 extended haplotypes  
 JOURNAL (in) Kasahara, M. (Ed.):  
 MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND  
 FUNCTION: 110-130;

REFERENCE 2 (bases 1 to 1074)  
 AUTHORS Ziegler, A., Ehlers, A., Forbes, S.A., Trowsdale, J., Volz, A.,  
 Younger, R.M., and Beck, S.  
 TITLE Polymorphisms in olfactory receptor genes: a cautionary note  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 1074)  
 AUTHORS Younger, R.M., Amado, C., Bethel, G., Ehlers, A., Fischer, L., K.,  
 Forbes, S.A., Horton, R., Milne, S., Mungall, A.J., Trowsdale, J.,  
 Volz, A., Ziegler, A., and Beck, S.  
 TITLE Characterisation of clustered MHC-linked olfactory receptor genes  
 JOURNAL Unpublished  
 REFERENCE 4 (bases 1 to 1074)  
 AUTHORS Ehlers, A., Beck, S., Forbes, S.A., Trowsdale, J., Volz, A., Younger, R.  
 and Ziegler, A.  
 TITLE MHC-linked olfactory receptor loci exhibit polymorphism and  
 contribute to extended HLA/OR-haplotypes  
 JOURNAL Unpublished  
 REFERENCE 5 (bases 1 to 1074)  
 AUTHORS Ehlers, A.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik,  
 Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,  
 GERMANY

FEATURES  
 source location/Qualifiers  
 1..1074  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /cell\_line="LG2"  
 1..1074  
 /gene="6M1-10\*01"  
 1..1074  
 /gene="6M1-10\*01"  
 /product="olfactory receptor"  
 /protein\_id="CAC20506.1"  
 /db\_xref="GI:12054397"  
 /translation="MNWYKSVPOEFILVSDPWFLEIPFVWFLSYIITITGNLT  
 IILVSHVDFKHTPTWYFPLSLSLDLCYTTSTVPMVLNCRKVIISVGGVQALF  
 IFLVSGTECLLAVMCFDRFVAICRPLHSIIMHRCFQLAASWISGFSNVLS  
 TWTLKMPICGHEVDHFECEVPALDKLSVDTANAELEFISVFLIPVTLILISY  
 AFIVQAVLRISAGORAKFCGSHLIVSLFPGTAISYLOPPSPSSKDRGKWSL  
 FCGIAPMLNPILYTLRNKEVKAERLVAKSLNQEIRNMQNSPFAKDYLTLYTNF  
 SASPIRVITLLENCLPQRKEP"

BASE COUNT 278 a 248 c 202 g 346 t  
 ORIGIN

alignment\_scores:  
 Quality: 1319.00 Length: 314  
 Ratio: 4.517 Gaps: 1  
 Percent Similarity: 92.994 Percent Identity: 82.484

alignment\_block:  
 US-09-755-017-2 x HSA302586 ..  
 Align seg 1/1 to: HSA302586 from: 1 to: 1074

1 MetAsnTrrPValAsnAspSerIleIleGlnGluPheIleLeuGlnLys 17  
 1 ATGAATTTGGTAATTAAGAGTGTCCACAGAGGATTCATTCTGTACTTT 50  
 17 eSerAspArgProTrrPLeuGluPhePheLeuValAlaPheLeuLys 34  
 51 CTCAGATCAACCATGCTAGAGATTCACCCCTTGGAGTGTCTGCTTTT 100  
 34 eTrrThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50  
 101 CCTATATCTTGACAAATCTTGGCAATCTGACATATATTTCTTGTCACAT 150  
 51 LeuAspThrLysLeuHisThrProMetTyrPhePheLeuThrAsnLeu 67



```

1 ATGATTTGGGTAAATAAGACTGTCCACAGAGCTCATCTGTAGTTT 50
17 eserAspArgProTyrLeuGluPheProLeuLeuValPheLeuIleS 34
51 CTAGATCAACCATGGCTAGAGATTCACCCCTTGTGATGTTCTGTTT 100
34 eTyrThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50
101 CCTATATCTTGACAACTCTTGGAATCTGCATAATTTCTGTGTACAT 150
51 LeuAspThrIleLeuIleThrProMetTyrPhePheLeuThrAsnLeuS 67
151 GTGATTTCAACACTCCACACCCCTATGCTATTTTCTTACCAATCTCTC 200
67 rLeuLeuAspLeuCysTyrThrThrCysThrValProGlnMetLeuVal 84
201 ACCCTGGAGCTTGTCTATACCAAGTACAGTTCACAAATCTGTGTAA 250
84 snLeuCysSerIleArgValIleSerTyrArgGlyCysValAlaGln 100
251 ACATATGCAACACACAGAAAGTATCATGTGTGTGTGTGTGTGTGTGT 300
101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuLeuAla 117
301 CTTTCAATTTCTGGCTGTGGCTTCCACAAATGTTCTTCTGTGTGTGT 350
117 lMetSerPheAspArgPheValAlaIleCysArgProLeuIleThrSer 134
351 CAGTGTCTTGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 400
134 aIleMetHisGlnArgLeuCysLeuGlnLeuAlaAlaAspTyrVal 150
401 TTATCATGACACAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 450
151 ThrGlyPheSerAsnSerValTyrPheSerThrLeuThrLeuGlnLeu 167
451 AGGGCTTTAGCAATTCAGTATTCAGTCCACCTGACACTTAAGATGCC 500
167 oleuCysAspProTyrValIleAspHisPheLeuCysGluValProAl 184
501 ACGTGTGTGTACAAAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 550
184 euLeuLeuLeuSerCysValGluThrThrAlaAsnGlnAlaGluLeu 200
551 TGTCTCAAGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 600
201 LeuValSerGluPhePheIleLeuProLeuThrLeuIleLeuIleS 217
601 TTCAATGAGTGTCTATCTCTTAATATACCCGTGACACTCATCTTAT 650
217 rTyrAlaPheIleValAlaGlyAlaValLeuArgIleGlnSerAlaG 234
651 GTATGCTTTTATTTGTCCAAAGAGTGTGTGTGTGTGTGTGTGTGTGT 700
234 rGlnIleValAlaPheGlyThrCysGlySerHisLeuIleValValS 250
701 AACGAAAGCATTTGGGACATGTGGCTCCCATTAATGTGTGTGTGTGT 750
251 PheTyrSerThrAlaValSerValTyrLeuGlnProProSerProSe 267
751 TTTTATGAGTACAGCTATCTCATATGACGTGCAACCCCTTCCAGCTC 800
267 rLysAspGlnGlyLysMetValSerLeuPheTyrGlyIleIleAlaP 284
801 CAAAGACCGGGGAAAGATGTTCTCTCTTCTGTGTGTGTGTGTGTGTGT 850
284 eLeuAsnProLeuIleTyrThrLeuArgAsnGlyValValLysGluG 300
851 TCTGTATTCCTCTTATATATACCTTAGGACAAAGAGTAAAGAGAGCC 900
301 PheLysArgLeuValAlaArgValPheLeuIleLysLys 313
901 TTTTAAAGGTTGTTGCAAAGAGTCTTCTTAATCAAGAA 940

```

```

seq_name: gb_pr: HSA302588
seq_documentation_block:
LOCUS HSA302588 1074 bp DNA PRI 06-JAN-2001
DEFINITION Homo sapiens 6M1-10*01 gene for olfactory receptor, cell line WT51.
ACCESSION AJ302588
VERSION AJ302588.1 GI:12054400
KEYWORDS 6M1-10*01 gene; olfactory receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J., Volz,A.,
uchanska,Ziegler,B., Volz,A., Younger,R. and Beck,S.
Polymorphic olfactory receptor genes and HLA loci constitute
extended haplotypes
(in) Kasahara,M. (Ed.);
MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND
FUNCTION: 110-130;
Springer-Verlag, Tokyo, Japan (2000)
2 (bases 1 to 1074)
Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J., Volz,A.,
Younger,R.M. and Beck,S.
Polymorphisms in olfactory receptor genes: a cautionary note
unpublished
3 (bases 1 to 1074)
Younger,R.M., Amadou,C., Bethel,G., Ehlers,A., Fischer,Lindahl,K.,
Forbes,S.A., Horton,R., Milne,S., Mungall,A.J., Trowsdale,J.,
Volz,A., Ziegler,A. and Beck,S.
Characterisation of clustered MHC-linked Olfactory Receptor Genes
in Human and Mouse
unpublished
4 (bases 1 to 1074)
Ehlers,A., Beck,S., Forbes,S.A., Trowsdale,J., Volz,A., Younger,R.
and Ziegler,A.
MHC-linked olfactory receptor loci exhibit polymorphism and
contribute to extended HLA/OR-haplotypes
unpublished
5 (bases 1 to 1074)
Ehlers,A.
Direct Submission
Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik,
Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,
GERMANY
FEATURES
source
1..1074 Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/cell_line="WT51"
1..1074
/gene="6M1-10*01"
1..1074
/gene="6M1-10*01"
/codon_start=1
/product="Olfactory receptor"
/protein_id="CAC20508.1"
/db_xref="GI:12054401"
/translacion="MNMVNSVPOEFILVFSDDPMLPEPFVWFILSYLTIGNLT
IIIVSHDFLHMPVPELISNLISLDICVTSYVPOMLVNICTRKYISVGGCAOLF
IFLAGSTEBCLLAVMCFDFRVALICRPLHRSIIMHQCQLAASWISGFSNSVLOS
TWIKMPLCGHKKEVDIFCEVPALIKLSVDVTANAELEFISVLFILPVTLILISY
AFIVQAVLRIOSEGGORAKRGTSGLIVSLVSGTAVISMYLQPPSPSSDRGKMWL
FCGLIAPMLPLATYTLNRNKEVKEAFKRLVAKSLINQDIRMNMWISFAKDVIVLTGTF
SASCPIVITITENVCNLPQRKFP"
BASE COUNT 278 a 248 c 202 g 346 t
ORIGIN

```

```

alignment_scores:
Quality: 1319.00 Length: 314

```

Ratio: 4.517 Gaps: 1  
Percent Similarity: 92.994 Percent Identity: 82.484

Alignment block:  
US-09-755-017-2 x HSA302588 ..

Align seg 1/1 to: HSA302588 from: 1 to: 1074

```

1 MetAspTrpValAsnAspSerIleIleGlnGluPheIleLeuGluGly 17
1 ATGAATGGGTAATAGAGAGTGTCCACAGAGAGTGTATCTTGTAGTTT 50
17 eSerAspArgProTrpLeuGluPheProLeuValValPheLeuIle 34
51 CTCAGATCAACCATGGCTAGAGATTCACCCCTTGTGATGTTCTTTT 100
34 eTrpThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50
101 CCTATATCTTGGACAAATCTTGGCAATCTGACATATATCTTGTGCACAT 150
51 LeuAspThrIleLeuHisThrProMetIlePhePheLeuThrAsnLeu 67
151 GGGATTTCAAACTCCACACCCCTATGACTTTTCTTTCAGCAATCTCTC 200
67 rLeuLeuAspLeuCysTrpThrCysThrValProGlnMetLeuValA 84
201 ACTCCGGACCTTGTGCTATACCAAGTACAGTCCACAAATGCTGTAA 250
84 snLeuCysSerIleArgIleValIleSerTrpArgGlyCysValAlaGln 100
251 ACATATGCAACACACAGAAAGTAATCAGTATGCTGTGCTGTGGCCAG 300
101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTrpLeuLeuValA 117
301 GTTTCATATCTTCCCTGGCTGTGGTCCACAGAAATGCTTCTCTGCGCG 350
117 lMetSerPheAspArgPheValAlaIleCysArgProLeuHisTrpSerV 134
351 CATGTGCTTGTAGTAGTGTGTAGCTATTTGTGCGCTCCCATTAATCA 400
134 aIleMetHisGlnArgLeuGlyLeuGlnLeuAlaAlaAspTrpVal 150
401 TTATCATGACACAGAGGCTGTGCTTCACAGTGTGACACTGCATCTGAT 450
151 ThrGlyPheSerAsnSerValTrpLeuSerThrLeuThrLeuGlnLeu 167
451 AGTGGCTTAGCAATTCAGTATTAACAGTCCAGCTGAGCACTTAAGTGC 500
167 oLeuCysAspProTrpValIleAspHisPheLeuGlyValProAla 184
501 ACTGTGTGTCACAAAGAGTGATCACTTCTGTGTAAGTCCCTGCTC 550
184 euleuLeuLeuSerCysValGluTrpThrAlaAsnGluAlaGluLeuPhe 200
551 TGGTCAAGTGTGCTGTGTGACACAAAGCAATGAGGCTGAATATTC 600
201 LeuValSerGluLeuPheHisLeuIleProLeuThrLeuIleLeuLeu 217
601 TTCATAGAGTGTCTATCTTCTTAATACCGGTGACACTCATCTTATATC 650
217 rTyraAlaPheIleValArgAlaValLeuArgIleGlnSerAlaGluGly 234
651 GTATGCTTTTATTTGTCCAGAGAGTGTGAAGATCCAGTCTGCGAAGTC 700
234 rGlnIleValAlaPheGlyThrCysGlySerHisLeuIleValValSerLeu 250
701 AACGAAAGGCATTTGGACATGTGGCTCCCATCTAATTGGGTGCTCAT 750
251 PheTrpSerThrAlaValSerValTrpLeuGlnProProSerProSer 267
751 TTTTATGGACAGCTATCTCAGTACGTGACCAACCTTCACCCAGCTTC 800
267 rLysAspGlnGlyLysMetValSerLeuPheTrpGlyIleIleAlaPro 284

```

```

|||||.....
801 CAAGACCGGGAAGATGTTCTCTCTCTGTGGAATCATGACACCA 850
284 eLeuAsnProLeuIleTrpThrLeuArgAsnLysGluValLysGluGly 300
851 TGGTGAATCCCTTATATATACACTTAGGAACAAAGGTAAAGGAAGCC 900
301 PheLysArgLeuValAlaArgValPheLeuIleLysLys 313
901 TTTAANAAGTTGGTTCAGAAAGAGTCTTCTTAATCAAGAAA 940

seq_name: gb_pr:HSA302589

seq_documentation_block: 1074 bp DNA PRI 06-JAN-2001
LOCUS HSA302589 1074 bp DNA PRI 06-JAN-2001
DEFINITION Homo sapiens 6M1-10*01 gene for olfactory receptor, cell line SA.
ACCESSION AJ302589
VERSION AJ302589.1 GI:12054402
KEYWORDS 6M1-10*01 gene; olfactory receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1074)
Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J.,
Uchanska-Ziegler,B., Volz,A., Younger,R. and Beck,S.
Polymorphic olfactory receptor genes and HLA loci constitute
extended haplotypes
(in) Kasahara,M. (Ed.):
MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND
FUNCTION: 110-130;
Springer-Verlag, Tokyo, Japan (2000)
2 (bases 1 to 1074)
Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J., Volz,A.,
Younger,R.M. and Beck,S.
Polymorphisms in olfactory receptor genes: a cautionary note
3 (bases 1 to 1074)
Younger,R.M., Amadou,C., Bethel,G., Ehlers,A., Fischer,Lindahl,K.,
Forbes,S.A., Horton,R., Milne,S., Mungall,A.J., Trowsdale,J.,
Volz,A., Ziegler,A. and Beck,S.
Characterisation of clustered MHC-linked olfactory Receptor Genes
in Human and Mouse
Unpublished
4 (bases 1 to 1074)
Ehlers,A., Beck,S., Forbes,S.A., Trowsdale,J., Volz,A., Younger,R.
and Ziegler,A.
MHC-linked olfactory receptor loci exhibit polymorphism and
contribute to extended HLA/OR-haplotypes
Unpublished
5 (bases 1 to 1074)
Ehlers,A.
Direct Submission
Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik,
Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,
GERMANY

FEATURES
Source
Location/Qualifiers
1..1074
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/cell_line="SA"
1..1074
/gene="6M1-10*01"
1..1074
/gene="6M1-10*01"
/codon_start=1
/product="Olfactory receptor"
/protein_id="CAC20509.1"
/db_xref="GI:12054403"
/translation="MNWVNRKSVPOEFLILVSDPWEIPFPWFLEFSYILITFENLT
IILVSHDFKLTPEYFELSNLSLDLCYTVSTVPMQVNTCRKRVISYSGCVAQLR
IFLAGSTECILLVWCPDRVAICRPLHYSIIMHQRICFOLAAASWISGFSNVLAGS

```

TWTLKMP LCGHEVDHFECEVPA LKLSVDVTINAE LFFISVLFPVTLILISY  
AFIQAVALRIQASBQKAFCTGCSHLIVASLEFGTATSM LOPSPSSKRGKAVSL  
FCGIAPLNP LITLRNKEVKEAPRLVARS LNLQEI RNNQMSFAKDVTULTYLTJNF  
SASCP IEFITL IENCNLPQRKFP"

BASE COUNT 278 a 248 c 202 g 346 t

ORIGIN

alignment\_scores:

Quality: 1319.00 Length: 314

Ratio: 4.517 Gaps: 1

Percent Similarity: 92.994 Percent Identity: 82.484

alignment\_block:

US-09-755-017-2 x HSA302589

Align seg 1/1 to: HSA302589 from: 1 to: 1074

1 MetAsnTrpValAsnAspSerIleIleGlnGluPheIleLeuLeuGly 17  
|||||  
1 ATGAATTGGGTAATAAGAGTGTCCACAGAGAGTTCATCTGTAGTTT 50  
17 eSerAspArgProTyrLeuGlnPheProLeuLeuValPheLeuIle 34  
|||||  
51 CTGAGATCAACACCTGCTAGAGATCCACCCCTTGTGATGTTCTGT 100  
34 ertYrThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50  
|||||  
101 CCTATCTTGTGACATCTTTGGCAATCTGCATTAATCTTGTCATCAT 150  
51 LeuAspThrLysLeuHisThrPrometLysPhePheLeuThrAsnLeu 67  
|||||  
151 GTGAGATTTCAAATCCACACCCCTATGATCTTTTCTTATGACATCTCTC 200  
67 rLeuLeuAspLeuGlyTyrThrThyCysThrValProGlnMetLeuVal 84  
|||||  
201 ACCCTGGAGCTTGTGCTATACACAGTACAGTTCACCAATCTCGTAA 250  
84 snLeuGlySerIleArgLysValIleSerTyrArgLysValAlaGln 100  
|||||  
251 ACATATGCAACACACAGAGAGTATGATGTTATGCTGTGCGCCAG 300  
101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuAlaVal 117  
|||||  
301 CTTTTCATTTTCTCGGCTTGGCTTCCACAGATGCTTCTCTCGGCCGT 350  
117 lMetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSer 134  
|||||  
351 CATGCTCTTGTATAGGTTGTACCTATTGTCGCGCTCTCCATTTACTCA 400  
134 alIleMetHisGlnArgLysLeuGlnLeuAlaAlaIleSerTyrVal 150  
|||||  
401 TTATCTATGCCACAGAGGCTCTGCTTCAGTGGACACTGATCTGATTT 450  
151 ThrGlnPheSerAsnSerValTyrLeuSerThrLeuThrLeuGlnLeuP 167  
|||||  
451 AGAGGCTTTAGCAATTCAGATTAAGTCCACCTGACACTTAAGATGCC 500  
167 oLeuGlyAspProTyrValIleAspHisPheLeuGlyValProAla 184  
|||||  
501 ACTGTGTGTCAACAAAGAGTGATGATCTTCTGTGAAGTCCCTGCTC 550  
184 euLeuLysLeuSerCysValGluThrThrAlaAsnGluAlaGluLeuPhe 200  
|||||  
551 TGTCTCAAGTTGCTCTGTGTTGACACACAGCAATAGGCTGACATTAATC 600  
601 TTCATCAGTGTGCTATTCTCTTAATACCGGTGACACTCATCTTATATC 650  
217 rTyrAlaPheIleValAlaValAlaLeuArgIleGlnSerAlaGluGly 234  
|||||  
651 GATATGCTTTATTTGTCCAAAGAGTGTGAGAAATCCAGTCTGCGAAGTTC 700

seq\_name: gp-pr:HSA302590

seq\_documentation\_block:

LOCUS HSA302590 1074 bp DNA PRI 06-JAN-2001

DEFINITION Homo sapiens 6M1-10\*01 gene for olfactory receptor, cell line YAR.

ACCESSION AJ302590.1 GI:12054404

VERSION AJ302590.1 GI:12054404

KEYWORDS 6M1-10\*01 gene; olfactory receptor.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

CDS

```

/gene="6M1-10*01"
1..1074
/gene="6M1-10*01"
/codon_start=1
/product="Olfactory receptor"
/protein_id="CAC20510.1"
/db_xref="GI:12054405"
/translation="MNNVNSKSVPOEFILVPSDOPMLEIPFVMEFLSYLTITFIGNLT
IIIVSHYDFKHTPMPEFLSNLSDLCYTSYTPQMLVNCNTRKYSYSGCYAOLF
ITLALSTBECLLVKCFDRVAICRPLHSIIMHQLCFOLASWISGFSNVSLOS
TWILKMPICGKHEVDHFCEVPALEKLSVDVTNEMAELEFISVFLIPVTLILISY
AFIVQAVLRLOSAGORAKFGTCSHLIVSLFYGTASIMYLQPPSPSSKRGKMWSL
FCGIIAPMLNPLIYLRNKVEKFAFKRLVAKSLNIOEIRNMQISPAKDFIVLYITNF
SASCPITVITLNCNLPQRKFP"
BASE COUNT      278 a      248 c      202 g      346 t
ORIGIN

```

## alignment\_scores:

Quality: 1319.00 Length: 314  
 Ratio: 4.517 Gaps: 1  
 Percent Similarity: 92.994 Percent Identity: 82.484

## alignment\_block:

US-09-755-017-2 x HSA302590

Align seg 1/1 to: HSA302590 from: 1 to: 1074

```

1 MetAsnTrpValAsnAspSerIleIleGlnGluPheIleLeuGlnGly 17
|||||
1 ATGAATGGGTAATTAAGAGTGTCCACAGAGAGTCAATCTGTGATTTT 50
17 eSerAspArgProTrpLeuGluPheProLeuLeuValAlaPheLeu 34
|||||
51 CTCAGATTCACACCTGGCTAGAGATTCACCTTTGTGATGTTCTGTATT 100
34 eTrpThrValThrIlePheGlyAsnLeuThrIleLeuValSerArg 50
|||||
101 CCGATATCTTGACAACTTTGGCAATCTGACATATTTCTGTGTACAT 150
51 LeuAspTrpLysLeuHisThrProMetTyrPhePheLeuThrAsnLeu 67
|||||
151 GTGGATTTGAACCTCCACACCCCTATGTAATTTTCTTATGCAATCTCTC 200
67 rLeuLeuAspLeuGlyTyrThrTrpCysThrValProGlnMetLeuVal 84
|||||
201 ACTCGTGGACCTTTGCTATACACAGTACAGTTCACAAATGCTGGTAA 250
84 snLeuGlySerIleArgLysValIleSerTyrArgLysValAlaGln 100
|||||
251 ACATATGCACACACGAGAAAGTAACTAGTATGCTGTGTGGCCAC 300
101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuAlaVal 117
|||||
301 CTTTATATTTTCTGGCTGTGGCTTCACAGATGTTCTCTCTGCGCGT 350
117 lMetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSer 134
|||||
351 CATGTCTTTGATAGGTTGTAGCTATTTGTCGCTCTCCATTAAGTCAA 400
134 alIleMetHisGlnArgLeuGlnLeuAlaIleAlaSerTrpVal 150
|||||
401 TTATCATGCACCGAGAGCTGCTCTCCAGTGGCAGCTGATCTGTGATT 450
151 ThrGlyPheSerAsnSerValTyrPheSerThrLeuThrLeuGlnLeu 167
|||||
451 AGTGGCTTTAGCAATTCAGATTACAGTCCACCTGACACTTAAGATGCC 500
167 oLeuGlyAspProTyrValIleAspHisPheLeuGlyValProAla 184
|||||
501 ACTGTGTGTGCACAAAGAGTGGATCATCTTCTGTGAAGTCCCTGCTC 550
184 euleuLysLeuSerCysValGluThrThrAlaAsnGluAlaGluPhe 200

```

seq\_name: gb\_pr: HSA302591

seq\_documentation\_block:

```

LOCUS      HSA302591      1074 bp      DNA      PRI      06-JAN-2001
DEFINITION Homo sapiens 6M1-10*01 gene for olfactory receptor, cell line OLGA.
ACCESSION  AJ302591
VERSION    AJ302591.1 GI:12054406
KEYWORDS   6M1-10*01 gene; olfactory receptor.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 1074)
            Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J., Volz,A.,
            Uchanska-Ziegler,B., Volz,A., Younger,R. and Beck,S.
            Polymorphic olfactory receptor genes and HLA loci constitute
            extended haplotypes
            (in) Kasahara,M. (Ed.):
            MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND
            FUNCTION: 110-130.
            Springer-Verlag, Tokyo, Japan (2000)
REFERENCE  2 (bases 1 to 1074)
            Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J., Volz,A.,
            Younger,R.M. and Beck,S.
            Polymorphs in olfactory receptor genes: a cautionary note
            Unpublished
            3 (bases 1 to 1074)
            Younger,R.M., Amadou,C., Bethel,G., Ehlers,A., Fischer Lindahl,K.,
            Forbes,S.A., Horton,R., Milne,S., Mungall,A.J., Trowsdale,J.,
            Volz,A., Ziegler,A. and Beck,S.
            Characterisation of clustered MHC-linked olfactory Receptor Genes
            in Human and Mouse
            Unpublished
            4 (bases 1 to 1074)
            Ehlers,A., Beck,S., Forbes,S.A., Trowsdale,J., Volz,A., Younger,R.
            and Ziegler,A.
            MHC-linked olfactory receptor loci exhibit polymorphism and
            contribute to extended HLA/OR-haplotypes
            Unpublished
            5 (bases 1 to 1074)
            Ehlers,A.
            Direct Submission

```

JOURNAL Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik,  
Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,

## FEATURES

GERMANY

source Location/Qualifiers

1..1074

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="6"

/cell\_line="OLGA"

1..1074

/gene="6M1-10\*01"

1..1074

/gene="6M1-10\*01"

/codon\_start=1

/product="Olfactory receptor"

/protein\_id="CAC20511.1"

/db\_xref="GI:12054407"

/translation="MNWVKSVPQEFLLVPSDFWLEIPFVWLFSLITTEGNT  
ILVSHVDFKLHTPMYFSLNLSLDLYTSVQMLNVCNTRKVISGCVQQLF  
IFLAGSTECFLIAMCFDRFAICRPHYSIIMHRCFOLAASWISGFSNVLOS  
TWTLKMPICGKREVDHFCEVPALIKISCDPTAEALFPLISVLLIPVLLISY  
APIQAVVRIOSAEGRKAEKSTGSHLVVSLFYSTAISMLQPPSPSKDRKMWL  
FCGIIAPLNPPLITLIRNKEVKAERKRLVANSLSLQETRNQMSFADVLITLTFN  
SASCPFIETIENCNLPORKEP"

BASE COUNT 278 a 248 c 202 g 346 t  
ORIGIN

## alignment\_scores:

Quality: 1319.00 Length: 314  
Ratio: 4.517 Gaps: 1  
Percent Similarity: 92.994 Percent Identity: 82.484

## alignment\_block:

US-09-755-017-2 x HSA302591 ..

Align seg 1/1 to: HSA302591 from: 1 to: 1074

```

1 MetAsnTrpValAsnAspSerIleIleGluPheIleLeuGlyPh 17
|||||
1 ATGAATGGGTAAATAAGAGTGTCCACAGAGAGTCTCTGTGTTAGTTT 50
17 eSerAspArgProTrpLeuGluPheProLeuLeuValPheLeuIles 34
|||||
51 CTCAGATCAACACGAGCTAGATTCACCCCTTGATGATGTTCTGTTT 100
34 eTrpTrpValIleIlePheGlyAsnLeuThrIleIleLeuValSerArg 50
|||||
101 CCTATATCTTGGACAAATCTTGGCAATCTGCACATAATCTTGTGCACAT 150
51 LeuAspThrLysLeuHisThrPrometTrpPhePheLeuThrAsnLeu 67
|||||
151 GTGGATTTCAAACTCCACACCCCTATGACTTTTCTTACCAATCTCTC 200
67 rLeuLeuAspLeuGlyTrpThrCysThrValProGlnMetLeuValA 84
|||||
201 ACTCGTGAGACCTTGTCTATACACAGTACAGTTCACAAATCTGTGTA 250
84 snLeuGlySerIleArgValIleSerTrpArgGlyCysValAlaGln 100
|||||
251 ACATATGCAACACGAGAAAGTATAGTATGTGTGTGTGTGTGTGTGTGT 300
101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTrpLeuLeuLeuAla 117
|||||
301 CTTTATATTTCTCGGCTTGGCTTGCACAGAAATGCTTCTCTGCGCGT 350
117 lMetSerPheAspArgPheValAlaIleCysArgProLeuHisTrpSer 134
|||||
351 CAGTGCTTTGATAGGTTTGTACTATTGTGCGCTCTCCATTAATCAAA 400
134 alIleMetHisGlnArgLeuGlyLeuGlnLeuAlaAlaSerTrpVal 150
|||||
401 TTATCATGACACGAGAGCTCTGCTTCCAGTGTGACAGCTGATCCTGATT 450

```

```

151 ThrGlyPheSerAsnSerValTrpLeuSerThrLeuThrLeuGlnLeuPr 167
|||||
451 AGTGGCTTTAGCAATTCAGATTCAGATTCACCTGACACTTAAGATGCC 500
167 OleuGlyAspArgProTrpValIleAspHisPheLeuGlyValAlaProAla 184
|||||
501 ACTGTGTGTGCACAAAGAGTGCATGCTCTGTGTAAGTCCCTGCTC 550
184 euleuGlyLeuSerGlyValGluThrThrAlaAsnGluAlaGluLeuPhe 200
|||||
551 TGCTCAAGTTGCTCTGTGTGACACACACCAATGAGGCTGACATATTC 600
201 LeuValSerGluLeuPheHisLeuIleProLeuThrLeuLeuLeuIle 217
|||||
601 TTATATCGTGTGCTATTTCTTAAATACCCGTGACACTCATCTTATATC 650
217 rTyraIlePheIleValAlaGluValLeuArgIleGlnSerAlaGluGlyA 234
|||||
651 GTATGCTTTATTTGTCACACGAGTGTGACAAATCCAGTCTGGAAGTTC 700
234 rGluIleValAlaPheGlyThrCysGlySerHisLeuIleValSerLeu 250
|||||
701 AACGAAAGGCAATTTGGACATGTGGCTCCCATCTAATTTGGTGTACTT 750
251 PheTrpSerThrAlaValSerValTrpLeuGlnProProSerProSer 267
|||||
751 TTTTATGTTAGTACACTATCTTCATGATCTGACCAACCTTCACCAAGCTC 800
267 rLysAspGlnGlyLysMetValSerLeuPheTrpGlyIleIleAlaProM 284
|||||
801 CAAGAAGCCGGGAAAGATGTTCTCTCTCTGTCGATCATTCACACCCA 850
284 eLeuAsnProLeuIleTrpThrLeuArgAsnLysGluValLysGluGly 300
|||||
851 TGCTGATCCCTTATATATATACCTTAGGAACAAAGAGTAAGAAAGACC 900
301 PheLysArgLeuValAlaArgValPheLeuIleLysLys 313
|||||
901 TTTTAAAGGTTGGTTGCCAAGAGTCTTCTTAATCAAGAAA 940
seq_name: gb_pr:HSA302592
seq_documentation_block:
LOCUS HSA302592 1074 bp DNA PRI 06-JAN-2001
DEFINITION Homo sapiens 6M1-10*01 gene for olfactory receptor, cell line AMAT.
ACCESSION AJ302592
VERSION AJ302592.1 GI:12054408
KEYWORDS 6M1-10*01 gene; olfactory receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 - (bases 1 to 1074)
Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J.,
Uchanska-Ziegler,B., Volz,A., Younger,R. and Beck,S.
Polymorphic olfactory receptor genes and HLA loci constitute
extended haplotypes
(in) Kasahara,M. (Ed.);
MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND
FUNCTION: 110-130;
Springer-Verlag, Tokyo, Japan (2000)
REFERENCE
2 (bases 1 to 1074)
Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J., Volz,A.,
Younger,R.M. and Beck,S.
Polymorphisms in olfactory receptor genes: a cautionary note
unpublished
3 (bases 1 to 1074)
Younger,R.M., Anadou,C., Bethel,G., Ehlers,A., Fischer,Lindahl,K.,
Forbes,S.A., Horton,R., Milne,S., Mungall,A.J., Trowsdale,J.,
Volz,A., Ziegler,A. and Beck,S.
Characterisation of clustered MHC-linked Olfactory Receptor Genes
in Human and Mouse

```



JOURNAL Unpublished  
 REFERENCE 4 (bases 1 to 1074)  
 AUTHORS Ehlers, A., Beck, S., Forbes, S.A., Trowsdale, J., Volz, A., Younger, R. and Ziegler, A.  
 TITLE MHC-linked olfactory receptor loci exhibit polymorphism and contribute to extended HLA/OR-haplotypes  
 JOURNAL Unpublished  
 REFERENCE 5 (bases 1 to 1074)  
 AUTHORS Ehlers, A.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik, Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin, GERMANY

FEATURES  
 source Location/Qualifiers  
 1..1074  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /cell\_line="AMA1"  
 1..1074  
 /gene="6M1-10\*01"  
 1..1074  
 /gene="6M1-10\*01"  
 /codon\_start=1  
 /product="Olfactory receptor"  
 /protein\_id="CAC20512.1"  
 /db\_xref="GI:12054409"  
 /translation="MNWVKSVPDEFILVFSDDPWEIIPPWFLEFSYILITFIGNLTILVSHDFELKHTPMYFELSLDLDCYTSYVQMLVNICNTRKVISYGCVAOLFIFLAGSTECFLAVMCFDRFAICRPLHYSIIMHREFOLAAWSISGFSNYSIOWTWLKMPLCGHREVDHFECEVDALIKSCVDPTANAELEFISVFLIPYLLIISYAFIVQAVLRIOASGQRACFGTCSHLIVSLFPGTASMTLOPSPSKRGKWSLFCGIAPMLNPLITLIRNKEAFKRLVAASLNLQETRMQMISFADTVLTITLNF SASCPITFITTENCNLPORRFP"

BASE COUNT 278 a 248 c 202 g 346 t  
 ORIGIN

alignment\_scores:  
 Quality: 1319.00 Length: 314  
 Ratio: 4.517 Gaps: 1  
 Percent Similarity: 92.994 Percent Identity: 82.484

alignment\_block:  
 US-09-755-017-2 x HSA302592 ..

Align seg 1/1 to: HSA302592 from: 1 to: 1074

```

1 MetAsnTrpValAsnAspSerIleIleGlnGluPheIleLeuGlyPh 17
|||||
1 ATGAATTGGTAAATTAAGACTGTCCACAGAGATTCTATCTGTAGTTT 50
17 eSerAspArgProTrpLeuGluPheProLeuLeuValPheLeuIles 34
|||||
51 CTCAGATCAACACATGGCTAGAGATCCACCTTGTGATTTCTGTTT 100
51 rIleuLeuAspLeuGlySerIleThrIleValPheGlnMetLeuVal 84
|||||
201 ACTCCGTGACCTTGGCTATACACAGATGATCCCAATGCTGTGTA 250
201 sLeuGlySerIleArgValIleSerIleValGlyValAlaGln 100
|||||
251 ACATATGCAACACACAGAAAGTATGATGCTGTGTCGCGCAG 300
251 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTrIleuLeuAlaVa 117

```

```

|||||
301 CTTTCAATTTCTCGGCTGGGTTCCACAGATGTCTTCTCGGCGCT 350
117 lMetSerPheAspArgPheValAlaIleCysArgProLeuIleIleSerV 134
|||||
351 CATGTGCTTGTAGATGTTGTAGCTATTTGTGCGGCTCTCATATACAA 400
134 aIleMetHisGlnArgLeuGlyLeuGlnLeuAlaAlaIleSerTrpVal 150
|||||
401 TTATCATGACACAGAGGCTCTGCTCCAGTTGGCAGCTGCATCCGATT 450
151 ThrGlyPheSerAsnSerValTrpLeuSerThrLeuThrLeuGlnLeuP 167
|||||
451 AGTGGCTTTAGCAATTCAGATTACAGTCCACCTGGACACTTAAGATGCC 500
167 oLeuGlyAspProTrpValIleAspHisPheLeuGlyValProAlaL 184
|||||
501 ACTGTGTGTCACAAAGAGTGGATGATCACTTCTGTGAGAGTCCGCTC 550
184 eLeuGlySerLeuGlyValGluThrThrAlaGlnGluAlaGluPhe 200
|||||
551 TGCTCAAGTTGCTCTGTGTGACACACAGCAATGAGGCTGAATATTC 600
201 LeuValSerGluLeuPheHisLeuIleProLeuThrLeuLeuIleSe 217
|||||
601 TTCAATCATGCTGCTATTCCTTCTAATACCCGTACACTCATCTTATATC 650
217 rTyraAlaPheIleValArgAlaValLeuArgIleGlnSerAlaGluGly 234
|||||
651 GTATGCTTTTATTTGTCACACAGAGTGAATATCCAGTCTGCGAAGGTC 700
234 rGlnGlnAlaIlePheGlyThrCysGlySerHisLeuIleValIleSerLeu 250
|||||
701 AACGAAAGGCAATTTGGACATGTGGCTCCCATTAATTTGGGTCACTT 750
251 PheTySerThrAlaValSerValTyLeuGlnProProSerProSerSe 267
|||||
751 TTTTATGATGACAGCTATCTCATGTACCTGCAACCACTTCACCCAGCTC 800
267 rLysAspGlnGlyLysMetValSerLeuPheTyGlyIleIleAlaProM 284
|||||
801 CAAGAAGCGGGGAAAGATGTTCTCTCTCTGTAATCATTTGACACCA 850
284 eLeuAsnProLeuIleIleTyThrLeuArgAsnLysGluValLysGluGly 300
|||||
851 TGCTGAATCCCTTATATATACACTTGAACAAGAGGTAAAGGAGGCC 900
301 PheLysArgLeuValAlaArgValPheLeuIleLysLys 313
|||||
901 TTTRAAAGGTTGGTTGCAAGAGCTCTTATATCAAGAAA 940
seq_name: gb_pr:HS193B12
seq_documentation_block:
LOCUS HS193B12 100375 bp DNA PRI 23-NOV-1999
DEFINITION Human DNA sequence from PAC 193B12 on chromosome 6p21.3-22.3.
Contains histone H2A/d, H2B/d, H2A.1, H1.5, H3.F, H4.k, H3.J
genes, histone ph2b.1 and hypothetical protein A4 pseudogenes,
histone H2A.1 and H2B.2 duplicate genes, Glycine (GGC) tRNA gene,
olfactory receptor OL1 like gene, Etsr srs and predicted Cpg
islands.
ACCESSION Z98744
VERSION Z98744.1 GI:3080457
KEYWORDS 6p21.3-22.3; A4; Cpg island; glycine tRNA; H1.5; H2A; H2A.1;
H2A/d; H2B; H2B.2; H2B/d; H3; H3.1; H3.F; H3/j; H4; H4.k; histone;
OL1; olfactory receptor; ph2b.1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 100375)
AUTHORS Wild, A.
TITLE Direct Submission

```

## JOURNAL

## COMMENT

Submitted (12-DEC-1997) Chromosome 6 Project Group  
 (http://www.sanger.ac.uk/HGP/Chr6/) Sanger Centre, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 On Apr 24, 1998 this sequence version replaced g1:2578067.  
**IMPORTANT:** This sequence is the entire insert of clone 193B12.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variations annotated may not be found in the sequence submission  
 corresponding to the overlapping clone as we submit sequences with  
 only a small overlap as described above.  
 This sequence was generated from part of bacterial clone contigs of  
 human chromosome 6, constructed in collaboration by the Sanger  
 Centre chromosome 6 mapping group and Armin Volz & Andreas Ziegler,  
 David Ruddy, Jeffrey Gruen.

Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr6/ This sequence has been finished  
 according to sequence map criteria as follows. An attempt is made  
 to resolve all sequencing problems, such as compressions and  
 repeats, but not necessarily within known annotated human repeat  
 sequence elements (e.g. Alu). Where the sequence is ambiguous,  
 there is an annotation using the 'unsure' feature key.  
 The true left end of clone 193B12 is at 100375.  
 The true right end of clone 193B12 is at 100375.  
 193B12 is from the library Rpci1 constructed at the Roswell Park  
 Cancer Institute by the group of Pieter de Jong.  
 For further details see http://bacpac.med.buffalo.edu/.

## FEATURES

## source

1. 100375  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /map="6p21.3-22.3"  
 /clone="Rpi-193B12"  
 /clone\_11b="Rpci-1"  
 211. 312  
 /note="MIR2 repeat: matches 33. 137 of consensus"  
 534. 768  
 /note="Alu repeat: matches 289. .35 of consensus;  
 incomplete repeat"

repeat\_region  
repeat\_region

## prln\_transcript

complement(1001. 1308)  
 /note="match: 5' EST H54606"  
 complement(<1589. .>2087)  
 /gene="H2A/d"

## mRNA

/note="match: ESTs AA426351 AA188780 AA15892 AA048695  
 AA061654 H15508 WA1535 AA272868 AA453105 W65215 AA231513  
 AA073767 AA528953 AA492640 W11262 AA060372 AA027719  
 AA143419 G90551; match: cDNAs E19778 L19778 X58069 X14850  
 D17284; match: genomic DNAs Z83739 X83549 Z83736 Z83742  
 X05862 Y00117 X16148 Z30940 U62674 X57138 U62669 X02218  
 U62673 X07763 X07758"  
 /evidence="not\_experimental"  
 /product="histone H2A"  
 /gene="H2A/d"  
 complement(1589. .2087)  
 /gene="H2A/d"  
 complement(1688. .2080)  
 /gene="H2A/d"  
 /codon\_start=1  
 /evidence="not\_experimental"  
 /product="histone H2A"  
 /protein\_id="CAB11417.1"  
 /db\_xref="GI:3080458"  
 /translation="MSGKGGKQKARAKAKTRSSRAGLPVGRVHRLLRGNTAERV  
 GAGAPVLAIVLEVTATLEAGNAARDNKRTIIPHLQLATINDELNLKLGKVT  
 IAQGVLPNIQAVLPKTESHKRKGK"  
 1807. 2816  
 /note="putative CPG island"

gene  
CDSmisc\_feature  
mRNA

<2327. .>2857  
 /gene="H2B/d"  
 /note="H2B.1 and H2B.2 like; match: ESTs AA354038 AA217829  
 WA9399 AA396588 AA080125 WA1523 AA218213 AA423665 AA562098  
 AA493023 AA237851 WA1584 AA255145 AA327803 AA064568"

## gene

## CDS

AA066051 W73546 AA493453 AA315768 AA280257 R76857 AA256738  
 NA2055 N99711 N31296 N31249 W71872 AA255154 N28495  
 AA218213 AA286933 R79451 WA1523 AA286932 NA1584 AA066051  
 W73546 AA280257 AA507431 AA577199 AA57205 AA010223 N29733  
 N22907 H81180 AA405358 AA354035; match: cDNAs M25487  
 M18045 M18046 M60756; match: genomic DNAs Z83336 Z83740  
 X02621 Z80779 Z80780 X57985 U62669 Z80782"  
 /evidence="not\_experimental"  
 /product="histone H2B"  
 2327. 2857  
 /gene="H2B/d"  
 2403. 2783  
 /gene="H2B/d"  
 /note="H2B.1' and H2B.2 like; match: proteins Q00715 P10854  
 P33778 P02278 P02294 Q64477 CE07075 P27795 CE05165 CE07280  
 P10853"  
 /codon\_start=1  
 /evidence="not\_experimental"  
 /product="histone H2B"  
 /protein\_id="CAB11418.1"  
 /db\_xref="GI:3080459"  
 /db\_xref="SPTREMBL:Q99877"  
 /translation="MPKSKAPAPKKGSKKAVTKAOKKDKKRRKRKESYSVYVK  
 VLKQVHPDTGISKAMGIMNSFVNDIFRIAGESRLAHYKRRSYTSREIOTAVRL  
 LPGLAKHAEVSEGTAKVTKYTSK"  
 2956. 3243  
 /note="Alu repeat: matches 13. 303 of consensus"  
 4365. 4499  
 /note="Alu repeat: matches 1. 134 of consensus;  
 incomplete repeat"  
 4501. 4556  
 /note="Alu repeat: matches 1. 134 of consensus;  
 incomplete repeat"  
 4786. 5085  
 /note="Alu repeat: matches 299. 1 of consensus"  
 6646. 6723  
 /note="Alu repeat: matches 299. 1 of consensus"  
 7038. 7077  
 /note="39 copies of 2 mer 82 & conserved"  
 /note="20 copies of 2 mer 85 & conserved"  
 7408. 7702  
 /note="Alu repeat: matches 3. 297 of consensus"  
 10347. 10644  
 /note="Alu repeat: matches 2. 300 of consensus"  
 11933. 11977  
 /note="MIR2 repeat: matches 100. 146 of consensus"  
 12051. 12094  
 /note="22 copies of 2 mer 98 & conserved"  
 12459. 12640  
 /note="Alu repeat: matches 120. 299 of consensus;  
 incomplete repeat"  
 12834. 13127  
 /note="Alu repeat: matches 1. 296 of consensus"  
 15893. 16409  
 /note="match: BAC end B47865"  
 17170. 17463  
 /note="Alu repeat: matches 293. 1 of consensus"  
 17841. 18149  
 /note="match: 5' ESTs D31045 D31402"  
 18070. 18215  
 /note="2 copies of 73 mer 95 & conserved"  
 18695. 18995  
 /note="Alu repeat: matches 1. 302 of consensus"  
 19116. 19178  
 /note="MIR repeat: matches 173. 110 of consensus"  
 20099. 20194  
 /note="MIR repeat: matches 1. 95 of consensus"  
 20197. 20490  
 /note="Alu repeat: matches 301. 8 of consensus"  
 20495. 20572  
 /note="MIR repeat: matches 84. 161 of consensus"  
 20574. 20881  
 /note="Alu repeat: matches 308. 1 of consensus"  
 20912. 20996  
 /note="FLAM\_C repeat: matches 1. 85 of consensus"

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

```

repeat_region 21030..21345
/Note="MT1C repeat: matches 169..466 of consensus"
repeat_region 22763..23053
/Note="MER2 repeat: matches 345..19 of consensus"
repeat_region 23261..23657
/Note="MT1F repeat: matches 480..44 of consensus"
repeat_region 23759..24062
/Note="AluSp repeat: matches 1..303 of consensus"
repeat_region 25062..25189
/Note="MIR2 repeat: matches 2..137 of consensus"
repeat_region 26058..26355
/Note="AluX repeat: matches 302..8 of consensus"
repeat_region 27721..27772
/Note="2 copies of 26 mer 96 % conserved"
gene complement(27803..28142)
/Note="27803..28142"
CDS complement(27803..28142)
/Note="pH2B.1"
/Note="pH2B.1"
/Note="match: proteins Q90859 P06899 P10854 P23527 P33778
O64524 P06900 P02279 P02281 P10853 O64477: match:
ESRs AA256738 AA327803 AA286933 AA315768 AA286932 AA280257
AA493453 AA423665: match: CDNAS M18046 M18045 M69756
M23487: match: genomic DNAs X83547 Z83336 Z80780 U62669
X59961 U77510 X57138"
/codon_start=1
/pseudo
evidence=not_experimental
misc_feature 27829..29369
/Note="putative Cpg island"
repeat_region 28190..28492
/Note="AluX repeat: matches 1..302 of consensus"
repeat_region 28508..28807
/Note="AluSg repeat: matches 2..303 of consensus"
mRNA <29070..>29531
/Note="H2A.1"

alignment_scores:
Quality: 1319.00 Length: 314
Ratio: 4.517 Gaps: 1
Percent Similarity: 92.994 Percent Identity: 82.484

alignment_block:
US-09-755-017-2 x HS193B12/rev
Align seg 1/1 to reverse of: HS193B12 from: 1 to: 100375

1 MetAsnTrpValAsnAspSerIleIleGluPheIleLeuGlyPhe 17
|||||
76061 ATGAATTGGTAAATTAAGAGTGTCCACAGGAGTTCCTGTGTTT 76012
17 eSerAspArgProTyrPheGluPheProLeuValAlaPheLeuIle 34
|||||
76011 CTCAGATTCAACATGCTAGAGATTCACCTTTGGATGTTCTGTTT 75962
34 eTyrValThrIlePheGlyAsnLeuThrIleLeuValSerArg 50
|||||
75961 CCTATATCTTGACATCTTTGGCAATCTGACATTAATCTTGTCACAT 75912
51 LeuAspThrIleLeuHisThrProMetTyrPhePheLeuThrAsnLeu 67
|||||
75911 GGGATTTCAAACCTCACACCCCTATGTTCTTTCTTAGCAATCTGC 75862
67 IleLeuAspLeuGlyTyrThrIleThrCysThrValProGlnMetLeuVal 84
|||||
75861 ACTCCGGACCTTTGCTATACACAGTACAGTTCACAAAGCTGTGTA 75812
84 snLeuCysSerIleArgValIleSerTyrArgIleCysValAlaGln 100
|||||
75811 ACATATGCACACACGAAAGTAAATGATTATGCTGTGTGCGCCAG 75762
101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuAlaVal 117
|||||
75761 CTTTTCATTTCTCGGCTTGGGTTCCACAGATGTTCTCTGCGCCGT 75712

```

```

117 lMetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSerV 134
|||||
75711 CATGTCCTTTGATAGAGTTGTAGCTATTGTGGGCTCTCCATTACTCA 75662
134 aIleMetHisGlnArgLeuCysLeuGlnLeuAlaIleAspTyrVal 150
|||||
75661 TTATCATGCACAGAGAGCTGCTGCTTCCAGTTGGCAGCTGCATCTG 75612
151 ThrGlyPheSerAsnSerValIlePheSerThrIleThrLeuGlnLeu 167
|||||
75611 AGGCGCTTAGCAATTCAGTATTCAGTACAGTCCACCTGACACTTAAG 75562
167 oleuCysAspProTyrValIleAspHisPheCysGluValProAla 184
|||||
75561 ACTGTGTGTCACAAAGAGTGATGCTTCTTGTGAAGTCCGTGCTC 75512
184 euleuLeuSerSerCysValGluThrThrAlaGlnIleValLeuPhe 200
|||||
75511 TGCTCAAGTTGCTCTGTGTGACACAAAGCAATGAGGCTGACTATTC 75462
201 leuValSerGluLeuPheHisLeuIleProLeuThrLeuIleLeu 217
|||||
75461 TTCATCAGTGTGCTATTCTCTTAATACCCGTGACACTCATCTATATC 75412
217 rTyrAlaPheIleValAlaValAlaLeuArgIleGlnSerAlaGluGly 234
|||||
75411 GTATGCTTTTATTTGTCACAGCACTGTGTGAGAAATCCAGTGTGAG 75362
234 rGlnIleValAlaPheGlyThrCysGlySerHisLeuIleValSerLeu 250
|||||
75361 AAGCAAGCATTTGGGACATGTGCTCCCATTAATGTGGTGTACTT 75312
251 PheTyrSerThrAlaValSerValTyrLeuGlnProPheSerProSer 267
|||||
75311 TTTTATGATGACACGATCTCCATCTGACGACCAACCACTTACCCAG 75262
267 rIysAspGlnGlyLysMetValSerLeuPheTyrGlyIleIleAlaPro 284
|||||
75261 CAAGACCGGGGAAAGATGTTCTCTCTGCTGTAATCATGACACCA 75212
284 eTleuAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGlu 300
|||||
75211 TGCTGATCCCTTATATATATACCTTAGACACAAAGTAAGGAAGCC 75162
301 PheLysArgLeuValAlaArgValPheLeuIleLysLys 313
|||||
75161 TTAAAGGTGGTGCACAAAGATCTTCTTAATCAAGAA 75122

seq_name: gb_hlg:AC024428

seq_documentation_block:
LOCUS AC024428 166758 bp DNA HTG 26-MAY-2000
DEFINITION Homo sapiens chromosome 3 clone RP11-59801 map 3, WORKING DRAFT
SEQUENCE 17 unordered pieces.
ACCESSION AC024428.3 GI:8076917
VERSION AC024428.3
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 166758)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 3, clone RP11-59801
JOURNAL Unpublished
2 (bases 1 to 166758)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhalil,B., Brown,A., Burkett,G., Campoliano,A., Castle,A.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeRellano,K., Dewar,K., Diaz,J.S., Dodgson,S., Domingo,M., Doyle,M.,
Fenster,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,

```

TITLE  
COMMENT

Galagan, J., Gardyna, S., Glnde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Largocque, K., Lehoczy, J., Levine, R., Lien, C., Liu, G., Locke, K., Macdonald, P., Marguis, N., McCarthy, M., McEwan, P., McGuire, A., McKernan, K., McPheters, R., Meldrum, J., Menus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirelli, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wymann, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 25, 2000 this sequence version replaced g1:727173.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

## Project Information

Center project name: L6934

Center clone name: 598\_O\_1

## Summary Statistics

Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 156318 bases at least Q40  
Consensus quality: 161224 bases at least Q20  
Consensus quality: 163451 bases at least Q20  
Insert size: 191000; agarose-ftp  
Insert size: 165158; agarose-ftp  
Quality coverage: 4.0 in Q20 bases; agarose-ftp  
Quality coverage: 4.6 in Q20 bases; sum-of-coverage

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 17 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```

1 1025: contig of 1025 bp in length
* 1026 1125: gap of 100 bp
* 1126 3025: contig of 1900 bp in length
* 3026 3125: gap of 100 bp
* 3126 4346: contig of 1221 bp in length
* 4347 4446: gap of 100 bp
* 4447 7420: contig of 2974 bp in length
* 7421 7520: gap of 100 bp
* 7521 10669: contig of 3149 bp in length
* 10670 10769: gap of 100 bp
* 10770 14258: contig of 3489 bp in length
* 14259 14358: gap of 100 bp
* 14359 18958: contig of 4600 bp in length
* 18959 19058: gap of 100 bp
* 19059 23968: contig of 4910 bp in length
* 23969 24068: gap of 100 bp
* 24069 28546: contig of 4478 bp in length
* 28547 28646: gap of 100 bp
* 28647 37679: contig of 9033 bp in length
* 37680 37779: gap of 100 bp
* 37780 45607: contig of 7828 bp in length
* 45608 45707: gap of 100 bp
* 45708 54439: contig of 8732 bp in length

```

```

* 54440 54539: gap of 100 bp
* 54540 63730: contig of 9191 bp in length
* 63731 63830: gap of 100 bp
* 63831 77842: contig of 14012 bp in length
* 77843 77942: gap of 100 bp
* 77943 98659: contig of 20717 bp in length
* 98660 98759: gap of 100 bp
* 98760 127843: contig of 29084 bp in length
* 127844 127943: gap of 100 bp
* 127944 166758: contig of 38615 bp in length.
Location/Qualifiers
1. 166758
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-59801"
/clone_lib="RP11 Human Male BAC"
1. 1025
/note="assembly_fragment"
1126. 3025
/note="assembly_fragment"
3126. 4346
/note="assembly_fragment"
clone_end:77
vector_side:left"
4447. 7420
/note="assembly_fragment"
7521. 10669
/note="assembly_fragment"
10770. 14258
/note="assembly_fragment"
14359. 18958
/note="assembly_fragment"
19059. 23968
/note="assembly_fragment"
24069. 28546
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
28647. 37679
/note="assembly_fragment"
37780. 45607
/note="assembly_fragment"
45708. 54439
/note="assembly_fragment"
54540. 63730
/note="assembly_fragment"
63831. 77842
/note="assembly_fragment"
77943. 98659
/note="assembly_fragment"
98760. 127843
/note="assembly_fragment"
127944. 166758
/note="assembly_fragment"

```

```

BASE COUNT 48778 a 33823 c 33960 g 48591 t 1606 others
ORIGIN

```

```

alignment_scores:
Quality: 1319.00 Length: 314
Ratio: 4.517 Gaps: 1
Percent Similarity: 92.994 Percent Identity: 82.484

```

```

alignment_block:
US-09-755-017-2 x AC024428

```

```

Align seg 1/1 to: AC024428 from: 1 to: 166758

```

```

1 MetasnrTpyValnsnspSerlleleGIngluPharlleleuLeuGlyph 17
|||||
99340 ATGAATTGGGTAAATAAGAGTGTCCACAGAGATTCATTCTGTACTTTT 99389

```

17	eSeRspRrPrProPrlaEuGLuPhPheLeuValPrlaEuILes	34
	.....	
99390	CTCAGATCAACCAATGGCTAGAGATTCCACCCTTTGGATGTTCTGTATT	99439
34	eTyrTrhValThrLlePheGLyaLLeuThrLleLleuValSerArg	50
	.....	
99440	CCATATATCTGCACATCTTTGGCAATCTGCACATATATCTGTGGTCACAT	99489
51	LeuAspTrhLysLeuHisThrPrometylTrpPhePheLeuThrAsnLeuSe	67
	.....	
99490	GTGGATTTTCAAACTCCACACCCCTATGTAATCTTTTCTTAGCAATCTCTC	99539
67	rLleuLysPrlaEuCysTyrTrhThrCysThrValProGlnMetLleuValA	84
	.....	
99540	ACTCTGGACCTTTTCTTACCCACAAAGTACAGTTCACCAAAATCTGGTAA	99589
84	sAlaEuCysSerLleArgLysValLleSerTyrArgGLyCysValaLagIn	100
	.....	
99590	ACATATGCACACACCCGAAAGTAATACAGTTATGTGGCTGTGGGCCACG	99639
101	LeuPheLlePrlaEuLysLleuGLuValaTrhGluTyrLleuLleuLysVala	117
	.....	
99640	CTTTTCATTTTCTGGCTGGCTGGTCCACAGAAATGTTCTTCTCTGGCCGT	99689
117	lMetSerPheAspArgPheValaLleCysArgPrlaEuHisTyrSer	134
	.....	
99690	CATGTGCTTTGATAGTTTGTATGATATTGTGGCCCTCCATATCTCA	99739
134	aLlMetMetHisGlnArgLeuCysLleuGlnLleuLysValaLleSerTrpVal	150
	.....	
99740	TTATCATGCACACAGAGGCTCTCTCCATCTGGACCTCATCTCTGGATT	99789
151	ThrGlyPheSerAsnSerValTrpLeuSerTrhLleuThrLleuGlnLeuTrp	167
	.....	
99790	AGTGGCTTTACCAATTCAGTATTAACAGTCCACTGGACACTTAAGAATGCC	99839
167	OLeucCysAspProTyrValLleAspHisPheLeuCysGluValProAlaL	184
	.....	
99840	ACTGTGTGGTCCAAAGAGAGTGAGTACACTTCTGTGAAGTCCCTGCTC	99889
184	euleLysLysLeuSerCysValGluTrhTrhAlaAsnGlnLysLysLeuPhe	201
	.....	
99890	TGCTCAAGTGTCTCTGTGTGACACACAGCAAAATAGAGCTGACATATTTC	99939
201	LeuValSerGluLeuPheHisLleuLlePrlaEuTrhLleuLleuLleSe	217
	.....	
99940	TTTCATCACTGTGCTATCTCTTCTAATACCCGACACATCATCTTATATC	99989
217	rTyrAlaPheLleValaLArgAlaValaLLeuArgLleGlnSerAlaGluGlyA	234
	.....	
99990	GTATGCTTTTATTTGTCCAAAGCGTGTGAGATCCAGTGTGCGGAAGTTC	100039
234	rGlnGlnLysAlaPheGluThrCysGlySerHisLleuLleValaLleuLys	250
	.....	
100040	AACGAAGGCAATTGGGACATGTGGCTCCCATCTAATTGTGGTCTACTT	100089
251	PheTyrSerTrhAlaValSerValLysLysLleuGlnProProSerProSer	267
	.....	
100090	TTTTATGTGTACAGTATCTTCCATGTACCTGCACCAACCTTTCACCCAGTTC	100139
267	rLysAspGlnGlyLysMetValSerLleuPheTyrGlyLleLleValaProm	284
	.....	
100140	CAAAAGACCGGGGAAGAATGGTTCTCTCTTCTGTGGAAATCATTTGCCACCA	100188
284	eLleuAsnProLeuLleTyrTrhLeuArgAsnLysGluValLysGluGly	300
	.....	
100190	TGCTGAATCCCTTATATATACACTTAGAACAAAGAGGTAAAGGAAGCC	100239
301	PheLysArgLysLeuValaLysArgValPheLeuLleLysLys	313
	.....	
100240	TTTAAAGGTGTGGTCCAAAGAGTCTTCTTATATCAAGAA	100279

seq_name:	gb:btg:AC025941
seq_documentation_block:	
LOCUS	AC025941 176277 bp DNA 26-MAY-2000
DEFINITION	Homo sapiens chromosome 6 clone RP11-635011 map 6, WORKING DRAFT
SEQUENCE	SEQUENCE, 15 unordered pieces.
ACCESSION	AC025941
VERSION	AC025941.2 GI:7408057
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (bases 1 to 176277)
TITLE	Barren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL	Homo sapiens chromosome 6, clone RP11-635011
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 176277)
	Barren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

TITLE  
JOURNAL  
COMMENT

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence\_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L8577
Center clone name: 635.0.11
-----
Summary Statistics
Sequencing vector: M13: M77815: 100% of reads
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.960731
Consensus quality: 168929 bases at least Q40
Consensus quality: 172827 bases at least Q30
Consensus quality: 174151 bases at least Q20
Insert size: 176000: agarose-fp
Insert size: 174877: sum-of-contigs
Quality coverage: 4.5 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated when the finished sequence
* as soon as it is available and the accession number will

```

\* as soon as it is available and the accession number will

	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="6"	
	/map="6"	
	/clone="RP11-635O11"	
	/clone_id="RPC1-11 Human	Male BAC"
	1..2997	
misc_feature	/note="assembly_fragment"	
	3098..5606	
	/note="assembly_fragment"	
	clone_end:SP6	
	vector_side:left"	
	5707..9209	
misc_feature	/note="assembly_fragment"	
	9310..13614	
misc_feature	/note="assembly_fragment"	
	13715..20639	
misc_feature	/note="assembly_fragment"	
	20740..26545	
misc_feature	/note="assembly_fragment"	
	clone_end:T7	
	vector_side:right"	
	26646..37068	
misc_feature	/note="assembly_fragment"	
	37169..48083	
misc_feature	/note="assembly_fragment"	
	48184..59015	
misc_feature	/note="assembly_fragment"	
	59116..70420	
misc_feature	/note="assembly_fragment"	
	70521..80466	
misc_feature	/note="assembly_fragment"	
	80567..92325	
misc_feature	/note="assembly_fragment"	
	92426..110572	
misc_feature	/note="assembly_fragment"	
	110673..131898	
misc_feature	/note="assembly_fragment"	
	131999..176377	
misc_feature	/note="assembly_fragment"	
DASE COUNT	54493 a 33357 c 33254 g 53772 t	1401 others

1	MetAsnTrpValAsnAspSerIleIleGlnGluPheIleLeuLeuGlyPh	17
41072	ATGAATTTGGTAAATAAGAGTGTCCACAGAGGATTCATTCTGTAGTTT	41122
17	eSerAspArgProTrpLeuGluPheProLeuValValPheLeuIle	34
41122	CTCAATCAACCATGGCGTAAGAAATTCACCCCTTGTGAGTTCTGTGTTT	41172
34	eTyThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg	50
41172	CCTAATCTTGACATCTTTGGCAATCTGCACAAATATTTCTTGTGCACAT	41222
51	LeuAspThrLysLeuIleHisThrProMetLysPhePheLeuThrAsnLeu	67
41222	GTGGATTTCAAACTCCACACACCCCTATGTAATCTTTTCTGACAAATCTCTC	41272
67	rLeuLeuAspLeuGlySerLysThrThrCysThrValProGlnMetLeuVal	84
41272	ACTCTGGACCTTTCTCTATACCAAGTACATCACTGTCCACAAATGCTGGTAA	41322
84	snLeuGlySerIleArgLysValIleSerTyArgGlyCysValAlaGln	101
41322	ACATATGCACACCCAGAAAGTAATCAAGTATAGGAGGCTGTGGCCCGAG	41372
101	LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuLeuAlaVal	117
41372	CTTTTCAATTTTCTGTGCGCTTGGGTGGTCCACAGAAATCTCTTCTGTGGCGCT	41422
117	IMetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSerV	134
41422	CATGGCTTTGTATAGTTTGTAGCTAATTTGTGGCCCTCTCCATTACTCA	147
134	AlleMetHisGlnArgLeuGlyCysLeuGlnLeuAlaAlaIleSerTyrVal	150
41472	TTATCATCTCACACAGAGGCTCTGCTCCAGTTGGAGCTGCATCTGGATT	41522
151	ThrGlyPheSerAsnSerValTyrLeuSerThrIleThrLeuGlnLeuPr	167
41522	AGTGAGCTTTAGCAATTCAGTATTATACAGTCCACCTGGACATTAACAATGCC	41572
167	OLEuGlyAspProTyrValIleAspHisPheLeuGlyGluValProAlaL	184
41572	ACTGTGTGTGCACAAAGAGTGCATCACTCTTCTGTGAAGTCCCTGCTC	41622
184	euleuLysLeuSerCysValGluThrThrAlaAsnGluAlaGluLeuPhe	200
41622	TGCTCAATGTTCCGTGTGTGCACACACAGCAAAAGAGGCTGACACTATTC	41672
201	LeuValIleSerGluLeuPheHisIleLeuProLeuThrIleLeuIleIle	217
41672	TTTCATCAAGTGTGCTATCTCTCTTAAATACCCGAGACATCAATCTTATATC	41722
217	rTyAlaPheIleValArgAlaValIleuArgIleGlnSerAlaGluGlyA	234
41722	GATATCTTTTATTTGTCCAAAGCAGTGTGAGAAATCAAGTCTGTGTAAGTCC	41772
234	rgGlnLysAlaPheGlyThrCysGlySerHisLeuIleValValSerLeu	250
41772	AACGAAAGCATTTGGAGCATGTGGCTCCCAATTAATTTGGTGTACATT	41822

```
251 PheTyrSerThrAlaValSerValTyrLeuGlnProProSerProSerSe 267
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
41822 TTTTATGGTACAGCTATCTCCATGTACCTGCACACACCTTCACCCAGCTC 41871
267 rLysAspGlnGlyLysMetValSerLeuPheTyrGlyIleIleAlaProM 284
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
41872 CAAAGACCGGGGAAAGATGGTTTCTCTCTCTCTGTCGAAATCATTCACCCA 41921
284 eLLeuAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGluGly 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
41922 TGCCTGAATCCCTTATATATACACTTAGGAACAAAGAGGTAAAGGAGAGCC 41971
301 PheLysArgLeuValAla.ArgValPheLeuIleLysLys 313
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
41972 TTTAAAGGTGTGTCAAAGAGCTCTCTTAATCAAGAAA 42011
```





OM of: US-09-755-017-2 to: N\_Geneseq\_1101.\* out\_format: pfs  
Date: Feb 26, 2002 5:28 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

```
-MODEL=framed.p2n.model -DEV=xlh  
-Q=c9n2_1/USFTO.spool/US09755017/runatc.25022002.165641.20808/app.query.fasta_1.374  
-DB=N_Geneseq_1101 -QFMT=fastseq -SUFFIX=p2n.rng -GAP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000  
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blonsum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct  
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODEL=LOCAL -OUTFMT=pfs  
-NORM_ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USRN=US09755017.@CGN1_1.213 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1
```

## Search information block:

```
Query: US-09-755-017-2  
Query length: 313  
Database: N_Geneseq_1101.*  
Database sequences: 930621  
Database length: 428662619  
Search time (sec): 112.140000
```

## score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation				
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				1601.00	2900.23	1.8e-153	939		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32019				1601.00	2900.19	1.8e-153	942		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32020				1601.00	2895.37	3.4e-153	1488		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32021				1319.00	2384.62	9.5e-125	1071		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32022				1319.00	2384.62	9.5e-125	1071		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32023				1138.00	2055.86	1.9e-106	939		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32024				1138.00	2051.44	3.4e-106	1442		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32025				1083.00	1959.59	4.5e-101	648		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32026				954.00	1720.01	9.9e-88	979		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32027				941.00	1702.35	9.6e-87	927		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32028				941.00	1696.81	1.9e-86	933		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32029				939.00	1693.37	3.1e-86	936		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32030				918.00	1654.02	4.7e-84	1011		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32031				912.00	1643.93	1.7e-83	936		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32032				908.00	1635.88	4.8e-83	1002		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32033				902.00	1619.82	3.8e-82	1629		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32034				898.50	1617.70	5.0e-82	1087		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32035				897.00	1616.54	5.8e-82	936		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32036				894.00	1610.35	1.3e-81	1002		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32037				891.00	1605.60	2.3e-81	936		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32038				889.00	1601.11	4.2e-81	1014		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32039				889.00	1597.41	6.7e-81	1440		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32040				883.00	1590.78	1.6e-80	957		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32041				883.00	1590.78	1.6e-80	957		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32042				882.00	1590.63	1.6e-80	917		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32043				879.00	1583.93	4.0e-80	954		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32044				863.50	1555.49	1.4e-78	933		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32045				861.00	1550.76	2.7e-78	948		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32046				861.00	1550.76	2.7e-78	948		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32047				859.00	1547.12	4.2e-78	948		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32048				853.00	1536.04	1.8e-77	960		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32049				849.00	1529.33	6.1e-77	909		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32050				810.50	1454.20	4.3e-73	1450		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32051				808.00	1455.78	5.2e-73	810		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32052				804.00	1446.30	1.7e-72	997		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32053				801.00	1441.29	3.3e-72	954		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32054				799.50	1438.06	4.7e-72	948		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32055				799.50	1437.71	5.3e-72	1000		
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32056				791.00	1433.43	3.3e-71	921		

/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32318 + 790.00 1421.05 4.5e-71 971  
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32319 + 787.00 1416.03 8.5e-71 930  
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32320 + 787.00 1415.79 8.7e-71 931  
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32321 + 786.50 1415.29 9.3e-71 915  
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32322 + 786.50 1415.26 9.4e-71 918

## seq\_name: /SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018

seq\_documentation\_block:  
ID AAH32018 standard; DNA; 939 BP.

AAH32018;

30-JUL-2001 (first entry)

Human olfactory receptor polynucleotide, SEQ ID NO: 591.

Human; olfactory receptor; OR; primary scent determination;

secondary scent determination; polypeptide library; odour receptor;

scent profile; scent fingerprint; scent representation; ds.

Homo sapiens.

MO200127158-A2.

19-APR-2001.

06-OCT-2000; 2000WO-US27582.

08-OCT-1999; 99US-0158615.

24-FEB-2000; 2000US-0184809.

(DIGI-) DIGISCENTS.

(YEDA ) YEDA RES & DEV CO LTD.

Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

WPI: 2001-290713/30.

New polynucleotides which encode polypeptides involved in olfactory

sensation for identifying olfactory agonists and antagonists -

Claim 8; Page 417; 1857pp; English.

The present sequence is one of a number of isolated polynucleotides

which encode polypeptides involved in olfactory sensation. The

polynucleotides can be used in screening for olfactory agonists and

antagonists. The methods allow for the determination of primary

scents and the identification of the odour receptors used to detect

these primary scents. The methods also enable determination of

secondary scents and the identification of combinations of odour

receptors that are involved in detecting such secondary scents.

This enables the construction of a scent representation (also called

a scent fingerprint or scent profile), which may be used to re-create

and edit scents. Libraries of olfactory receptors are useful for

determining the interaction pattern of a composition with the receptors,

and can be used for determining differences in the olfactory faculties

of different individuals.

Sequence 939 BP; 220 A; 227 C; 187 G; 305 T; 0 other;

## alignment\_scores:

Quality: 1601.00 Length: 313  
Ratio: 5.115 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-755-017-2 x AAH32018 ..  
Align seg 1/1 to: AAH32018 from: 1 to: 939

1 MetasnrTrpValAsnAspSerIleIleGlnGluPheIleuLeuGlyPh 17

```

|||||
1 ATGAATTGGGTAATGACACATCATACAGAGATTATCTGCGGGTTT 50
17 eSerAspArgProTirPleuGluPheProLeuValValPheLeuIles 34
51 CTGAGATCGACCTTGCGCTGGAGTTTCACCTCTTGCTCTTGATTT 100
34 eTyrThrValAlaThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50
101 CTACACTGTGACCATCTTTGGCATCTGACCATTTATCTAGTGTACGCG 150
51 LeuAspThrIleLeuHisThrProMetTyrPhePheLeuThrAsnLeuSe 67
151 CTGGACACCAACTTCATACCCCATGTATTTTCTTACCAATCTATC 200
67 rLeuLeuAspLeuGlyTyrThrThrCysThrValProGlnMetLeuValA 84
201 ACTGCTGGATCTTTGTACACACAGTGTACAGTCCCAATGCTACTAA 250
84 snLeuGlySerIleArgLysValIleSerTyrArgGlyCysValAlaGln 100
251 ATTATTCACACATCAGAAAGTATCATGTTATCGTGGCTGTAGAGCCAG 300
101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuLeuAla 117
301 CTTTTCATATTTCTGGCCTTGGGGCTACGAAATATCTTCTCCGCGCT 350
117 MetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSerV 134
351 CATGCTCTTGATAGCTTTGTACTATTTGTGCGCTCTCATTTACACAG 400
134 alIleMetHisGlnArgLeuGlyCysLeuGlnLeuAlaAlaSerTrpVal 150
401 TTAATCATGACACAGAGACTCTGCTCCAGTTGGCAGCTGATCTGGGTT 450
151 ThrGlyPheSerAsnSerValTirPleuSerThrLeuThrLeuGlnLeuPr 167
451 ACTGGTTTGTAGTAACCTAGTGTGTGTCTACACCTGACTCTCCAGCGCC 500
167 olauCysAspProTyrValIleAspHisPheLeuGlyValProAlaI 184
501 ACCTGTGACCCCTATGTGTATGATATCATCTTCTGTGAGAGTCCCTGAC 550
184 euLeuLysLeuSerCysValGluThrThraAlaAsnGluAlaGluLeuPhe 200
551 TGTCTAAAGTATCTTGTGTGTGAGACAAACAGCAATGAGCTGAACATTC 600
201 LeuValSerGluLeuPheHisLeuIleProLeuThrLeuIleLeuIleSe 217
601 CTGTGAGTGTAGCTCTTCCATTAATACCCCTGACACTCATCTTATATC 650
217 rTyrAlaPheIleValAlaArgAlaValLeuArgIleGlnSerAlaGluGly 234
651 ATATGCTTTATTTGTCGAGACACTATTGAGGATACACTGCTGTAAGGTC 700
234 rglInLysAlaPheGlyThrCysGlySerHisLeuIleValSerLeu 250
701 GACAAAGCATTTGGACATGTGGTCCCATCTAATTTGGGTGTCCTT 750
251 PheTyrSerThrAlaValSerValTyrLeuGlnProProSerProSerSe 267
751 TTTTATAGTACAGCCGCTCTGTGTACCTGCAACACACCTTGCCGACGCTC 800
267 rLysAspGlnGlyLysMetValSerLeuPheTyrGlyIleIleAlaProm 284
801 CAAGGACCAAGAAAGATGTTCTCTCTTCTATGGATATTGACACCA 850
284 eLeuAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGluGly 300
851 TGCTGATATCCCTTATATATACACTTAGAACAAGAGGTAAAGGAAGGC 900
301 PheLysArgLeuValAlaArgValPheLeuIleLysLys 313
|||||

```

```

901 TTTAAAGGTTGGTTCACAGAGCTTCTTATATCAAGAAA 939
seq_name: /SIDS2/gcgsdata/geneseq/geneseqn/AA2001.DAT:AA508541
seq_documentation_block:
ID AA508541 standard; cDNA; 942 BP.
XX
AC AA508541;
XX
DI 26-SEP-2001 (first entry)
XX
DE DNA encoding novel human G-protein coupled receptor (NGPCR).
XX
KW Human; novel G-protein coupled receptor; NGPCR; drug discovery;
XX diagnostic; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
ET 1.942
CDS /*tag= a
FT /product= "Novel G-protein coupled receptor (NGPCR)."
XX
XX WO200151634-A1.
XX
PD 19-JUL-2001.
XX
PE 05-JAN-2001; 2001WO-US00589.
XX
PR 12-JAN-2000; 2000US-0175764.
XX
XX (LEXI-) LEXICON GENETICS INC.
XX
PA Walke DW, Wilganowski NL, Turner CA, Friedrich G, Abulin A;
PI Zambrowicz B, Sands AT;
PI WPI; 2001-442145/47.
XX
DR P-PSDB; AAU04689.
XX
PT New human G protein-coupled receptor and polynucleotides encoding the
PT receptor, useful in identifying, selecting or validating new molecular
PT targets for drug discovery and in diagnostic or prognostic assays
XX
XX Claim 1; Page 61; 65pp; English.
XX
XX The sequence represents the coding sequence of novel human G-protein
XX coupled receptor (NGPCR). The NGPCR nucleic acid and protein are useful
XX in identifying, selecting or validating novel molecular targets for drug
XX discovery, as well as in diagnostic or prognostic assays. These are also
XX useful in microarrays or other assay formats, for screening collections
XX of genetic material from patients who have a particular medical condition
XX or for identifying mutations associated with a particular disease.
XX
SQ Sequence 942 BP; 222 A; 228 C; 187 G; 305 T; 0 other:

alignment_scores:
Quality: 1601.00 Length: 313
Ratio: 5.115 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-755-017-2 x AA508541
Align seg 1/1 to: AA508541 from: 1 to: 942
1 MetAsnTyrValAsnAspSerIleIleGlnGluPheIleLeuLeuGlyP 17
|||||
1 ATGAATTGGGTAATGACACATCATACAGAGATTATCTGCTGCGGTTT 50
17 eSerAspArgProTirPleuGluPheProLeuValValPheLeuIles 34
|||||
51 CTGAGATCGACCTTGCGCTGGAGTTTCACCTCTTGCTCTTGATTT 100
51 CTGAGATCGACCTTGCGCTGGAGTTTCACCTCTTGCTCTTGATTT 100

```

```

34  eTyrThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50
    |||||
101  CTTACACTGTGACCATCTTGGCAATCTGACATTAATTCAGTGCACGC 150
    |||||
51  LeuAspThrLysLeuHisThrProMetLysPhePheLeuThrAsnLeuSe 67
    |||||
151  CTGGACACCAAACTTCATACCCCATGATATTTTCTTACCAATCTATTC 200
    |||||
67  rIleuLeuAspLeuCysThrThrCysThrValProGlnMetLeuValA 84
    |||||
201  ACTCTGGATCTTGTACACCATGATACAGTCCCAAAATCTAGTAA 250
    |||||
84  snLeuCysSerIleArgLysValIleSerLysArgLysCysValAlaGln 100
    |||||
251  ATTATAGCAGCATCAGAAAGATATCATGTTATCGTGTGTGAGCCGAC 300
    |||||
101  LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuLeuAla 117
    |||||
301  CTTTCATATTTCTGGCTGGGCTGCTGATGATATCTTCTCGGCCGCT 350
    |||||
117  lMetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSerV 134
    |||||
351  CATGTCTTTGATGAGTTGTGTACATTTGTGCGCTCTCCATTTACTCAG 400
    |||||
134  alIleMetHisGlnArgLeuCysLeuGlnLeuAlaAlaSerTyrPval 150
    |||||
401  TTTATCATGACACGACAGACTCTGCTCCAGTTGGACGCCCATCTGGGTT 450
    |||||
151  ThrGlyPheSerAsnSerValThrLeuSerThrLeuThrLeuGlnLeuP 167
    |||||
451  ACTGGTTTATGTAACCTGAGTGTGTGTCTACCTGACTCTCCAGCTGCC 500
    |||||
167  OleuCysAspProTyrValIleAspHisPheLeuCysGluValProAla 184
    |||||
501  ACCTCTGTGACCCCTATGTATGATGATCACTTCTGTGAAGTCCCTGCAC 550
    |||||
184  euleuLysLeuSerCysValGluThrThrAlaAsnGluAlaGluLeuPhe 200
    |||||
551  TGTCTCAAGTTATCTTGTGTGGACACACGACAAATGAGCTGACATATTC 600
    |||||
201  LeuValSerGluLeuPheHisLeuIleProLeuThrLeuIleLeuIleSe 217
    |||||
601  CTGTGTGAGAGCTCTTCCATTAATACCCCTGACACTCATCTTATATTC 650
    |||||
217  rTyrAlaPheIleValAlaArgAlaValLeuArgIleGlnSerAlaGluGly 234
    |||||
651  ATATGCTTTTATTTGTCGAGACATATTGAGATACAGTCTGTGAAGGTC 700
    |||||
234  rGlnLysAlaPheGlyThrCysGlySerHisLeuIleValIleValSerLeu 250
    |||||
701  GACAAAAGCATTTGGGACATGTGTTCCCATTAATGTGGGTGCTCTT 750
    |||||
251  PheTyrSerThrAlaValSerValTyrLeuGlnProProSerProSerSe 267
    |||||
751  TTTTATATGATACGCCGCTCTGTGTACTGACCAACCACTTCCGCCAGCTC 800
    |||||
267  rLysAspGlnGlyLysMetValSerLeuPheTyrGlyIleIleAlaProM 284
    |||||
801  CAAGGACCAAGAAAGATGCTTCTCTTCTATGGAATCATTTGCAACCCA 850
    |||||
284  eLeuAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGluGly 300
    |||||
851  TGTCTGATCCCTTATATATATACCTTAAGACAAAGGAGTAAAGGAGAGGC 900
    |||||
301  PheLysArgLeuValAlaArgValPheLeuIleLysLys 313
    |||||
901  TTTAAAAGTGTGGTGCACAGAGTCTTCTTATCAAGAAA 939
    |||||
seq_name: /SID52/gcgdata/geneseq/geneseqn/NA2001.DAT: AAS08542
seq_documentation_block:
ID AAS08542 standard; cDNA: 1488 BP.

```

```

XX  AAS08542:
AC  26-SEP-2001 (first entry)
DT
DE  ORF encoding novel human G-protein coupled receptor (NGPCR).
KW  Human; novel G-protein coupled receptor; NGPCR; drug discovery;
    diagnostic; ss.
XX  Homo sapiens.
OS
XX  Key
FH  Location/Qualifiers
FT  198..1139
FT  CDS
FT  /tag=a
FT  /product="Novel G-protein coupled receptor (NGPCR)"

XX  WO200151634-A1.
XX  19-JUL-2001.
XX
XX  05-JAN-2001: 2001WO-US00589.
XX
XX  12-JAN-2000: 2000US-0175764.
XX
XX  (LEXI-) LEXICON GENETICS INC.
XX
XX  Walke DW, Wilganowski NL, Turner CA, Friedrich G, Abuin A;
XX  Zambrowicz B, Sands AT;
XX  WPI; 2001-442145/47.
XX  P-PSDB; AAU04689.
XX
XX  New human G protein-coupled receptor and polynucleotides encoding the
XX  receptor, useful in identifying, selecting or validating new molecular
XX  targets for drug discovery and in diagnostic or prognostic assays
XX
XX  Disclosure: Page 62-63; 65pp; English.
XX
XX  The sequence represents the open reading frame (ORF) encoding novel human
XX  G-protein coupled receptor (NGPCR). The NGPCR nucleic acid and protein
XX  are useful in identifying, selecting or validating novel molecular
XX  targets for drug discovery, as well as in diagnostic or prognostic
XX  assays. These are also useful in microarrays or other assay formats, for
XX  screening collections of genetic material from patients who have a
XX  particular medical condition or for identifying mutations associated with
XX  a particular disease.
XX
XX  Sequence 1488 BP: 421 A; 333 C; 254 G; 477 T; 3 other:

alignment_scores:
    Quality: 1601.00      Length: 313
    Ratio: 5.115          Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-755-017-2 x AAS08542 ..
Align seg 1/1 to: AAS08542 from: 1 to: 1488

1  MetAspTyrValAsnAspSerIleIleGlnGluPheIleLeuGlyP 17
    |||||
198  ATGAATTGGGTAAATGACACATCATACAGAGCTTATTTCTGCTGGGTTT 247
    |||||
17  eSerAspArgProTyrLeuGlnLubPheProLeuValAlaPheLeuL 34
    |||||
248  CTTGACATGCACCTTGGCTGAGATTTCACATCCCTTGGTCTTGTATTT 297
    |||||
34  eTyrThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50
    |||||
298  CTTACACTGTGACCATCTTGGCAATCTGACATTAATTCAGTGCACGC 347
    |||||

```

```

51 LeuAspThrIleuHisThrPrometTyrPhePheLeuThrAsnLeuSe 67
   |||||
348 CTGGACACCAACTTCATACCCCATGATATTTTCTTACCATATC 397
   |||||
67 rLeuLeuAspLeuCysTyrThrThrCysThrValProGluMetLeuValA 84
   |||||
398 ACTCTGGATCTTGTTCACACCATGTACAGTCCCAAAATGCTAGTAA 447
   |||||
84 snLeuCysSerIleArgLysValIleSerTyrArgGlyCysValAlaGln 100
   |||||
448 ATTATATGCACATCAGAAAGTATCAGTTATGCTGTGTGTAGTACCCAG 497
   |||||
101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuLeuAlaVal 117
   |||||
498 CTTTTCATATTTCTGGCCCTTGGGGGCTACTGAATATCTTCTCCGGCCGT 547
   |||||
117 lMetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSerV 134
   |||||
548 CATGTCTTGTGATAGGTTTGTACTATTTCTCGCCCTCTCCATTTACTCAG 597
   |||||
134 allMetHisGlnArgLeuCysLeuGlnLeuAlaAlaIleSerTrpVal 150
   |||||
598 TTATCATGCACACAGACACTCTGCTCCAGTTGGCAGCCGCACTCGGTT 647
   |||||
151 ThrGlyPheSerAsnSerValTrpLeuSerThrLeuThrLeuGlnLeuPr 167
   |||||
648 ACGGTGTTTGTAGTACAGTGTGGTGTCTACCCCTGACTCTCCAGCTGCC 697
   |||||
167 OlLeuCysAspProTyrValIleAspHisPheLeuCysGluValProAla 184
   |||||
698 ACCTCTGAGCCCTATCTGTAGATCCTTCTCTGGAAGTCCCTGCAC 747
   |||||
184 euleuLysLeuSerCysValGluThrThrAlaAsnGluAlaGluLeuPhe 200
   |||||
748 TGCCTCAAGTTATCTTGTGTGAGACAAACACCAATGAGGCTGACTATTC 797
   |||||
201 LeuValSerGluLeuPheHisLeuIleProLeuThrLeuIleLeuIleSe 217
   |||||
798 CTGTGAGTGTGAGCTCTCCATTAATACCCCTGACTCATTCCTTATATC 847
   |||||
217 rTyrAlaPheIleValAlaArgAlaValLeuArgIleGlnSerAlaGluGly 234
   |||||
848 ATATGCTTTTATTTGTCGACAGCATTTGAGGATACAGTCTGCTGAAGTCTC 897
   |||||
234 rGlnLysAlaPheGlyThrCysGlySerHisLeuIleValIleSerLeu 250
   |||||
898 GACAAAAGCATTTGGGACATGTGTTCCCATCTAATTTGGTGTCTCTT 947
   |||||
251 PheTyrSerThrAlaValSerValTyrLeuGlnProProSerProSerSe 267
   |||||
948 TTTTATATGTAAGCCGCTCTGTGTACCTGCAACACACCTTCCGCCACCTC 997
   |||||
267 rLysAspGlnGlyLysMetValSerLeuPheTyrGlyIleIleAlaProm 284
   |||||
998 CAAGGACCAAGAAAGATGTTCTCTCTATGGAATATATGCAACCA 1047
   |||||
284 etLeuAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGluGly 300
   |||||
1048 TGCCTGAATCCCTTATATATACCTTATGGAACAAAGGAGTAAAGGAGGC 1097
   |||||
301 PheLysArgLeuValAlaArgValPheLeuIleLysLys 313
   |||||
1098 TTTAAAGGTTGGTGCAGAGCTTCTTATATCAAGAAA 1136
   |||||
seq_name: /SID2/gcdata/geneseq/geneseqn/NA2001.DAT:AAH31669
seq_documentation_block:
ID AAH31669 standard; DNA; 1071 BP.
XX AAH31669;
AC
XX
XX 30-JUL-2001 (first entry)
XX

```

```

DE Human olfactory receptor polynucleotide, SEQ ID NO: 242.
XX
XX
KW Human; olfactory receptor; OR; primary scent determination;
KW secondary scent determination; polypeptide library; odour receptor;
KW scent profile; scent fingerprint; scent representation; ds.
OS Homo sapiens.
PN WO200127158-A2.
XX
XX 19-APR-2001.
XX
XX 06-OCT-2000; 2000WO-US27582.
XX
XX 08-OCT-1999; 99US-0158615.
XX
XX 24-FEB-2000; 2000US-0184809.
XX
XX (DIGI-) DIGISCENTS.
XX (YEDA ) YEDA RES & DEV CO LTD.
XX
XX Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
XX WPI, 2001-290713/30.
XX
XX New polynucleotides which encode polypeptides involved in olfactory
XX sensation for identifying olfactory agonists and antagonists -
XX
XX Claim 8; Page 273-274; 1857pp; English.
XX
XX The present sequence is one of a number of isolated polynucleotides
XX which encode polypeptides involved in olfactory sensation. The
XX polynucleotides can be used in screening for olfactory agonists and
XX antagonists. The methods allow for the determination of primary
XX scents and the identification of the odour receptors used to detect
XX these primary scents. The methods also enable determination of
XX secondary scents and the identification of combinations of odour
XX receptors that are involved in detecting such secondary scents.
XX This enables the construction of a scent representation (also called
XX a scent fingerprint or scent profile), which may be used to re-create
XX and edit scents. Libraries of olfactory receptors are useful for
XX determining the interaction pattern of a composition with the receptors,
XX and can be used for determining differences in the olfactory faculties
XX of different individuals.
XX
XX Sequence 1071 BP; 277 A; 248 C; 201 G; 345 T; 0 other.
XX
XX
XX alignment_scores:
XX      Quality: 1319.00      Length: 314
XX      Ratio: 4.517          Gaps: 1
XX      Percent Similarity: 92.994      Percent Identity: 82.484
XX
XX alignment_block:
XX US-09-755-017-2 x AAH31669 ..
XX
XX Align seg 1/1 to: AAH31669 from: 1 to: 1071
XX
XX 1 MetAsnTrpValAsnAspSerIleIleGlnGluPheIleLeuGluLys 17
XX |||||
XX 1 ATGAATTTGGGTAAATGAAGTCTCCCAAGGAGTTCATCTGTACTTTT 50
XX |||||
XX 17 eSerAspArgProTrpLeuGlnPheProLeuLeuValIlePheLeuIle 34
XX |||||
XX 51 CTCAGATCAACCACTGCTAGAGATTCACCCCTTGTGATGTTCTCTTTT 100
XX |||||
XX 34 eTyrThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50
XX |||||
XX 101 CCGATATCTTGACAAATCTTTGGCAATCTGACATATTTCTGTGTACAT 150
XX |||||
XX 51 LeuAspThrLysLeuHisThrPrometTyrPhePheLeuThrAsnLeuSe 67
XX |||||
XX 151 GTGATTTCAAACTCCACACCCCTATGTACTTTTCTTATGCAATCTCTC 200
XX |||||

```



84 snleuCySerlleArglyValIleSerTyArglyCysValAlaGln 100  
 101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyrIleuLeuAlaVa 117  
 301 CTTTTCATTTTCCTGGCCCTGGGGTCCACAGAAATGCTCTCTCCGGCCCT 350  
 117 lmetSerPheAspArgPheValAlaIleCysArgProLeuHISTySerV 134  
 351 CATGTGCTTGTATAGATGTTTGTAGCTATTTCGGCCCTCTCATTTACTCAA 400  
 134 alileuethISglArglyLeuCysLeuGlnLeuAlaAlaSerTyPval 150  
 401 TTTATCATGACACAGAGGCTCTGCTTCAGTTGGCAGCTGCATCTGGATTC 450  
 151 ThrGlyPheSerAsnSerValITrPLeuSerThrLeuThrIleuGlnLeuPr 167  
 451 AGTGGCTTTAGCAATTCAGATTACAGCTCCACCTGGACACTTAAGATGCC 500  
 167 oleuCyAspProTyValIleAspHisPheLeuCysGluValProAlaL 184  
 501 ACTGTGGTGCACAAAGAACTGCATCTCTCTGTGAAGTCCCTGCTC 550  
 184 euleuTyLeuSerCysValGluThrThraIlaasnGluAlaGluLeuPhe 200  
 551 TGCTCAAGTTGCTCCTGTGTGACACAAACAGCAATGAGCGCTGAACATTC 600  
 201 LeuValSerIleuLeuPheHisIleuIleProleuThrIleuIleuIleSe 217  
 601 TTTATCATGCTGTCTATTCTTTCAATAACCGTGACACTCATCTTAATTC 650  
 217 rTyValaPheIleValaArgAlaValaLeuAqIleGlnSerAlaGluGly 234  
 651 GTATGCTTTATTTGTCCACAGCACTGTGACAAATCCACTGCTGAAAGTCT 700  
 234 rglInuysAlaPheGlyThrCysGlySerHisIleuIleValSerLeu 250  
 701 AACGAAAGGATTTGGGACATGGGCTCCCATCTAATTTGGTGGTCACTT 750  
 251 PheTySerThrAlaValSerValTyrLeuGlnProProSerProSerSe 267  
 751 TTTATGAGTACACTATCTCATGTACTCTCAACACACCTTCACCCACTC 800  
 267 rTyAspGlnGlyIlysmetValSerIleuPheTyGlyIleIleAlaProM 284  
 801 CAAGAGACGGGGAAGATGCTTCTCTCTGTGGAAATCATTTGCACCCA 850  
 284 etleuAsnProLeuIleTyThrLeuArgAsnLysGluValLysGluGly 300  
 851 TGGTGAATCCCTTATATATACACTTAGAACAAGAGGTAAAGGAAGGCC 900  
 301 PheIlyArgIleuValAlaArgValPheLeuIleLysLys 313  
 901 TTTAAAGGTTGGTTCGCAAGAGCTCTTCAATCAAGAAA 940  
 seq\_name: /SID57/gcdata/geneseq/geneseqn/NA2001.DAT:AAH32304  
 seq\_documentation\_block:  
 ID AAH32304 standard; DNA: 939 BP.  
 XX AAH32304;  
 AC  
 XX  
 DT 30-JUL-2001 (first entry)  
 XX  
 DE Human olfactory receptor polynucleotide, SEQ ID NO: 877.  
 XX  
 KW Human: olfactory receptor; OR: primary scent determination;  
 XX secondary scent determination; polypeptide library; odour receptor;  
 KW scent profile; scent fingerprint; scent representation; ds.  
 XX  
 OS Homo sapiens.  
 XX

PN WO200127158-A2.  
 PD 19-APR-2001.  
 XX  
 PF 06-OCT-2000; 2000WO-0527582.  
 XX  
 PR 08-OCT-1999; 99US-0158615.  
 PR 24-FEB-2000; 2000US-0184809.  
 XX  
 PA (DIGI-) DIGISCENTS.  
 PA (YEDA) YEDA RES & DEV CO LTD.  
 XX  
 PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;  
 DR WPI: 2001-290713/30.  
 XX  
 PT New polynucleotides which encode polypeptides involved in olfactory  
 PS sensation for identifying olfactory agonists and antagonists -  
 XX  
 PS Claim 8: Page 530; 1857pp; English.

The present sequence is one of a number of isolated polynucleotides  
 which encode polypeptides involved in olfactory sensation. The  
 polynucleotides can be used in screening for olfactory agonists and  
 antagonists. The methods allow for the determination of primary  
 scents and the identification of the odour receptors used to detect  
 these primary scents. The methods also enable determination of  
 secondary scents and the identification of combinations of odour  
 receptors that are involved in detecting such secondary scents.  
 This enables the construction of a scent representation (also called  
 a scent fingerprint or scent profile), which may be used to re-create  
 CC and edit scents. Libraries of olfactory receptors are useful for  
 CC determining the interaction pattern of a composition with the receptors,  
 CC and can be used for determining differences in the olfactory faculties  
 of different individuals.  
 CC  
 XX

Sequence 939 BP; 232 A; 228 C; 185 G; 294 T; 0 other;

alignment\_scores:  
 Quality: 1138.00 Length: 313  
 Ratio: 4.138 Gaps: 0  
 Percent Similarity: 87.859 Percent Identity: 70.607

alignment\_block:  
 US-09-755-017-2 x AAH32304 ..

Align seg 1/1 to: AAH32304 from: 1 to: 939

1 MetAsnTrpValAsnSPSerIleIleGlnGluPheIleuLeuGlyP 17  
 1 ATCAATTTGGGAAATGAGAGCTCCCAAAAGACTTTATCTACTTGGCTT 50  
 17 eSerAspArgProTrPLeuGlnuPheProLeuLeuValaIlePheLys 34  
 51 CTCAGATAGGCGCTTGGCTACAAATGCCCTTTTGGTGGCTCTTAATAT 100  
 34 eTyTyThrValIThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50  
 101 CATACACAAATCACCATATTGGCAATGTGTCATCATGAGGTGCATT 150  
 51 LeuAspThrLysLeuHisThrPrometTyRPhPheLeuThrAsnLeu 67  
 151 CTGGATCCCAAACTTCATCTCCCATGATTTTCTCTCATATCTCTC 200  
 67 rIleuAspLeuCysTyThrThrCysThrValProGlnmetLeuVala 84  
 201 CATCTTAGATCTCTGCTATACCAACACTACAGCCCTCATATGTGGTAA 250  
 84 snleuCySerlleArglyValIleSerTyArglyCysValAlaGln 100  
 251 ATATTGGTTCGCAACAAAGAGCCATCATGCTATGCTGTGGCCAC 300

```

XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
XX thrombosis; contraceptive; ss.
OS Homo sapiens.
XX
XX W0200058473-A2.
PN
XX
XX 05-OCT-2000.
PD
XX
PF 31-MAR-2000; 2000OWO-US08621.
PR
XX 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000OUS-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Teach M;
PI
XX WPI: 2000-602362/57.
XX P-PSDB: AAB43266.
DR
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 5245-5246; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antiproliferative; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticoagulant; antifibrillic; immunosuppressant;
CC immunostimulant; cardiac; thrombolytic; coagulant; vasotrophic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antifibromatous; antibacterial; antiviral; antifungal; antineumatic;
CC antihypoid; and antinaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal hemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 1442 BP; 378 A; 319 C; 268 G; 477 T; 0 other;
SO

```

```

34 eTyrThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50
   |||.....|||.....|||.....|||.....|||.....|||.....|||
401 CATACAAATACACCATATTTGGCAATGTCCATCATGATGAGGTGCTATT 450
51 LeuAspThrLysLeuHisThrPrometTyrPhePheLeuThrAsnLeuSe 67
   |||.....|||.....|||.....|||.....|||.....|||.....|||
451 CTGGATCCCAAACTTCATCTCCATGATATTTCTTCTCTCACTATCTCTC 500
67 rLeuLeuAspLeuCysTyrThrThrCysThrValProGlnMetLeuVal 84
   |||.....|||.....|||.....|||.....|||.....|||.....|||
501 CATCTTAGATCTGCTATACACAACTACAGTCCCTCATATGTGTGTA 550
84 snLeuCysSerIleArgLysValIleSerTyrArgLysCysValAlaGln 100
   |||.....|||.....|||.....|||.....|||.....|||.....|||
551 ATATTGGTTCCAAACAAACATCATGCTATGCTGTGGCTGTGGCCAC 600
101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuAlaVal 117
   |||.....|||.....|||.....|||.....|||.....|||.....|||
601 CTCATCATCTTCCGTGGCCCTAGGTGCTACAGAGTCTCTCTGTGGCTGT 650
117 MetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSerV 134
   |||.....|||.....|||.....|||.....|||.....|||.....|||
651 TATGTCTTGTGACAGATATGTGGCTGTTCAGACCCCTCCACTATGTAG 700
134 alIleMetHisGlnArgLeuCysLeuGlnLeuAlaAlaIleSerTyrVal 150
   |||.....|||.....|||.....|||.....|||.....|||.....|||
701 TCATCATGATTTATGTGTTCTGCTAAGATGACAGCTTCTCATGTGCTC 750
151 ThrGlyPheSerAsnSerValTyrPheSerThrLeuThrLeuGlnLeuPr 167
   |||.....|||.....|||.....|||.....|||.....|||.....|||
751 ATTGGTTTCGCAACTGACGTGCTGCTCTTCTGCTGCTTCTTAACTGCC 800
167 oleuCysAspProTyrValIleAspHisPheLeuCysGluValProAlaL 184
   |||.....|||.....|||.....|||.....|||.....|||.....|||
801 ACCCTGTGTCACACGAAAGTGGACCACTTTCTGTGAGCTGCTGCAC 850
184 euleuLysLeuSerCysValGluThrThrAlaAsnGluAlaGluLeuPhe 200
   |||.....|||.....|||.....|||.....|||.....|||.....|||
851 TTCTCAAGTTGTCATGTGTCGACACAAAGCTATTGAGGCTGAGCTTCTC 900
201 LeuValSerGluLeuPheHisLeuIleProLeuThrLeuLeuLeuIleSe 217
   |||.....|||.....|||.....|||.....|||.....|||.....|||
901 TTCTTAGTCTACTAATTTCTTAAATTCCTGACATGATGATCTCATCTCTC 950
217 rTyrAlaPheIleValArgAlaValLeuArgIleGlnSerAlaGluGlyA 234
   |||.....|||.....|||.....|||.....|||.....|||.....|||
951 CTATGGCTTCATAGCTCAACGACTATTAATAATCAGGTGACGAGAGGAC 1000
234 rglInLysAlaPheGlyThrCysGlySerHisLeuIleValValSerLeu 250
   |||.....|||.....|||.....|||.....|||.....|||.....|||
1001 GGCATAAAGCATTTGGGACATGTGGGTCCACATGATGTGGTGTCTCTC 1050
251 PheTyrSerThrAlaValSerValTyrLeuGlnProProSerProSerSe 267
   |||.....|||.....|||.....|||.....|||.....|||.....|||
1051 TTTTATGGAACACGCACTTATATATCTTCAACACCACTTCATCCACCTTC 1100
267 rLysAspGlnGlyMetValSerLeuPheTyrGlyIleIleAlaProM 284
   |||.....|||.....|||.....|||.....|||.....|||.....|||
1101 TAAAGCACTGGGAAAGATGTTCCCTCTTCTATGGAATCATCATCACTCA 1150
284 eLeuAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGluGly 300
   |||.....|||.....|||.....|||.....|||.....|||.....|||
1151 TGTGTAAGTCTCCATCATCTACAGCCTTAAGAAATTAAGATNAGAAGGAG 1200
301 PheLysArgLeuValAlaArgValPheLeuIleLysLys 313
   |||.....|||.....|||.....|||.....|||.....|||.....|||
1201 TTCAAGAGGCTGATGCCAAGAAATCTTTTCTGTAAAGAAA 1239
seq_name: /SID52/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH32231
seq_documentation_block:
ID AAH32231 standard: DNA: 648 BP.

```

```

XX AAH32231;
AC 30-JUL-2001 (first entry)
DE Human olfactory receptor polynucleotide, SEQ ID NO: 804.
KW Human; olfactory receptor; OR; primary scent determination;
KW secondary scent determination; polypeptide library; odour receptor;
KW scent profile; scent fingerprint; scent representation; ds.
OS Homo sapiens.
PN WO200127158-A2.
PD 19-APR-2001.
XX 06-OCT-2000; 2000WO-US27582.
XX 08-OCT-1999; 99US-0158615.
PR 24-FEB-2000; 2000US-0184809.
XX (DIGI-) DIGISCENTS.
PA (YEDA ) YEDA RES & DEV CO LTD.
XX
XX Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
XX WPI: 2001-290713/30.
XX
XX New polynucleotides which encode polypeptides involved in olfactory
XX sensation for identifying olfactory agonists and antagonists -
XX
XX Claim 8: Page 502; 1857pp; English.
XX
XX The present sequence is one of a number of isolated polynucleotides
XX which encode polypeptides involved in olfactory sensation. The
XX polynucleotides can be used in screening for olfactory agonists and
XX antagonists. The methods allow for the determination of primary
XX scents and the identification of the odour receptors used to detect
XX these primary scents. The methods also enable determination of
XX secondary scents and the identification of combinations of odour
XX receptors that are involved in detecting such secondary scents.
XX This enables the construction of a scent representation (also called
XX a scent fingerprint or scent profile), which may be used to re-create
XX and edit scents. Libraries of olfactory receptors are useful for
XX determining the interaction pattern of a composition with the receptors,
XX and can be used for determining differences in the olfactory faculties
XX of different individuals.
XX
XX Sequence 648 BP; 144 A; 166 C; 132 G; 206 T; 0 other;
XX
XX
XX Alignment_scores:
XX      Quality: 1083.00      Length: 216
XX      Ratio: 5.061      Gaps: 0
XX Percent Similarity: 99.074      Percent Identity: 99.074
XX
XX alignment_block:
XX US-09-755-017-2 x AAH32231 ..
XX
XX Align seg 1/1 to: AAH32231 from: 1 to: 648
XX
XX 68 LeuLeuAspLeuCysTyrThrThrCysThrValProGlnMetLeuValAs 84
XX |||.....|||.....|||.....|||.....|||.....|||.....|||
XX 1 CTCTGGATCTTGTTCACACCATGTACAGTCCCAAAATGCTACTATAA 50
XX |||.....|||.....|||.....|||.....|||.....|||.....|||
XX nLeuCysSerIleArgLysValIleSerTyrArgLysCysValAlaGlnL 101
XX |||.....|||.....|||.....|||.....|||.....|||.....|||
XX 51 TTTATCAGCATCAGGAAAGTATATCATGCTATGCTGTGAGCCAC 100
XX |||.....|||.....|||.....|||.....|||.....|||.....|||
XX 101 eulPheIlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuAlaVal 117
XX |||.....|||.....|||.....|||.....|||.....|||.....|||
XX 101 TTTTCATATTTTTCGCTTGGGGGCTACGAAATATCTCTCGCGCGCTC 150

```



[illegible]

[illegible]

```
seq_name: /SDS2/gcgydata/geneseq/geneseqn/NA2001.DAT:AAH32067
seq_documentation_block:
ID   AAH32067 standard; DNA; 927 BP.
```

AAH32067;  
30-JUL-2001 (first entry)

Human olfactory receptor polynucleotide, SEQ ID NO: 640

Human; olfactory receptor; OR; primary scent determination; secondary scent determination; polypeptide library; odour receptor; scent profile; scent fingerprint; scent representation; ds.

**Homo sapiens.**

W0200127158-A2

19-APR-2001.

06-OCT-2000; 2000WO-US27582.

08-OCT-1999; 99US-0158615.

(YEDA ) YEDA RES & DEV CO LTD.

Belenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

WPI; 2001-290713/30.

New polynucleotides which encode polypeptides involved in olfactory sensation for identifying olfactory agonists and antagonists -

PS Claim 8; Page 437; 1857pp; English

The present sequence is one of a number of isolated polynucleotides which encode polypeptides involved in olfactory sensation. The polynucleotides can be used in screening for olfactory agonists and antagonists. The methods allow for the determination of primary scents and the identification of the odour receptors used to detect these primary scents. The methods also enable determination of secondary scents and the identification of combinations of odour receptors that are involved in detecting such secondary scents. This enables the construction of a scent representation (also called a scent fingerprint or scent profile), which may be used to re-create and edit scents. Libraries of olfactory receptors are useful for determining the interaction pattern of a composition with the receptors and can be used for determining differences in the olfactory faculties of different individuals.

50 Sequence 927 BP; 209 A; 253 C; 176 G; 289 T; 0 other

	alignment_scores:	
	Quality:	944.00
	Ratio:	3.688
	Percent Similarity:	84.768
		Length: 3022
		Gaps: 0
		Percent Identity: 59.603

alignment\_block:

Align seg 1/1 to: AAH32067 from: 1 to: 927

5 AsnAspSerIleIleGlnGlnIlePheIleLeuLeuGlyPheSerAspArgPr 21

13 AATGAGAGTTCCTAATGGAATTCATCCTTCTAGGCTTCTCAGACCACCC 62

21 OTRPLEUGLUPHEPROLEUVALPHELEULESERTYRTHRVALT 38

63 TCGCTGGAGGCTGTTCTCTTTGTATTTGTCCCTTTCTTCTTACCTCCTGA 112

38 hrlephgIyAsnLeuthrilelleuValSerArgleuAspThrLys 54

113 CCCTGTGGAACTTCACCATAATCATCATCTCATATCTGGATCCCCCT 162

55 LeuHisThrProMetTyrPhePheLeuThrAsnLeuSerLeuLeuAspLeu 71

163 CTTCAATACCCCAATGTA CTTTTTCTCAGCAACCTCTCTTACTGGACAT 21

71 ucystyrthrCysthrvalProGlnmetLeuValAsnLeuCysseri 88

213 CTGCTTCACTACTAGCCCTTGCTCCTCAGACCTTAGTTACTTGC AAGAC 262

88 leargLysvalIleSerTyrargGlycysvalAlaGlnLeupheIlephe 104

263 CAAAGAGACGATCATTACGGTGGTGTGTGGCCAACTCTATATTTCT 312

105 LeuAlaLeuGlyAlaThrGluTyrLeuLeuLeuAlaValMetSerPheAs 121

313 CTGGCACTGGGCTCCACTGAATGTATCCTCTTGGCTGACATGGCCTTGA 362

121 pargpHevalAlaIleCysArgProLeuHisTyrSerValIleMetHisG 138

363 TCGGTACATTGCTGCTGCAAAACCCCTCCACTATGTAGTCATCATGAACC 412

138 InArgLeuGlyLeuGlnLeuAlaAlaAlaSerTrpValThrGlyPheSer 154

413 CACGGCTTGGCAACAGCTGGCATCTATCTCCCTGGCTCAGTGGTTGGCT 462

155 AsnSerValTrpLeuSerThrLeuThrLeuGlnLeuProLeuGlyCysAspPr 171

463 AGTTCCTAATCCATGCCAACTTTTACCTTGCAATTGCCCTCTCTGTGGCAA 512

171 ctyvaalIleasprhlsPheleucysglIvalProalaleuleulysLeus 1888

513 CCATAGGCTGGACCATTTTATTTGCGAGTACCAGCTCTTCTCAAGTTGG 562

seq\_name: /SIDS2/gcdata/geneseq/geneseqn/NA2001.DAT:AAH32308

```
seq_documentation_block:
ID    AAH32308 standard; DNA; 933 BP.
```

AC ААНЗ2308;

DT 30-JUL-2001 (first entry)

Human olfactory receptor polynucleotide, SEQ ID NO: 881.

KW Human; olfactory receptor; OR; primary scent determination;

KW scent profile; scent fingerprint; scent representation; ds.

Homo sapiens.

PN W0200127158-A2.

PD 19-APR-2001.

PF 06-OCT-2000; 2000WO-US27582.

PR 08-OCT-1999; 99US-0158615.

XX  
XX  
G  
T  
O  
T  
C  
E  
R  
N  
A  
M  
E

PA (YEDA ) YEDA RES & DEV CO LTD.

PI Belenson J, Smith D, Glusman G, Fuchs T, Yanai I;  
 22

DR WPI; 2001-290713/30.

**PT** New polynucleotides which encode polypeptides involved in olfactory

XX  
PC  
CJ 91E 01. Dec 80 E31-E33. 105733. E31:ch

CC The present sequence is one of a number of isolated polynucleotides  
CC which encode polypeptides involved in olfactory sensation. The  
CC

CC polynucleotides can be used in screening for olfactory agonists and  
CC antagonists. The methods allow for the determination of primary  
CC scents and the identification of the odour receptors used to detect  
CC these primary scents. The methods also enable determination of  
CC secondary scents and the identification of combinations of odour  
CC receptors that are involved in detecting such secondary scents.  
CC This enables the construction of a scent representation (also called  
CC a scent fingerprint or scent profile), which may be used to re-create  
CC and edit scents. Libraries of olfactory receptors are useful for  
CC determining the interaction pattern of a composition with the receptors  
CC and can be used for determining differences in the olfactory faculties  
CC of different individuals.

50 Sequence 933 BP; 205 A; 239 C; 187 G; 302 T; 0 other.

```

alignment_scores:
  Quality: 941.00      Length: 302
  Ratio: 3.734        Gaps: 0
  Percent Similarity: 83.444  Percent Identity: 60.265

```

```
alignment_block:
US-09-755-017-2 x AAH32308 .
```

Align seg 1/1 to: AAH32308 From: 1 to: 933

4 ValAsnAspSerIleIleGlnGluPheIleLeuLeuGlyPheSerAsp 20

19 GTCAATGCTAGCTCTGAGGGTACTTTATTTTACGTGCAATTTCTAATTG 68

20 gprotRpleuGlupheproleuValValpheuileSerTyrrhv 37

69 GCCTCATCTGGAAGTACTTATCTTTGTGGTTGCTTGGATCTTCTACTTGA II

37 aI Thr Ile Phe Cys Asn Leu Thr Ile Ile Leu Val Ser Arg Leu Asp Thr 53

119 TGACACTGATAGGAACCTGTTCATCATCATCCTGTCAATACCTGGACTCC 16

54 LysLeuHisThrProMetTyrPhePheLeuThrAsnLeuSerLeuLeuAs 70

169 CATCTGCACACACCAATGTACTTCTTCCCTTCAACCTCTGATTTCTGGA 21

70 pleucystyrthrvalproglmetleuvalasnleucyss 87

219 TCTCTGCTACACCACCAGCTCTATCCCTCAGTTGCTGGTCAATCTCTGGG 26

87 erilearglyvalilesertyrargglycysvalalacnleupheile 10

269 GCCCGAAGAACCATCTCTATGCTGTTGCATGATTCACCTTACTTT 31

104 pheuuaIaLeuGIyAlaThrGIuTyrlEuLeuuaIaValmetSerph 12

319 GTTCTCGCACTGGGAACCAAGAGTGTCTACTGGTGGTGATGTCTTA 36

```
120 easpargpheValAlaIleCysargProLeuHisTyrSerValIleMeth 13
```

369 TGACCGTTATGCACGCTGTGTAGACCTTTGCATTACACTGTCCCTCATGC 41

137 1SG1NARGLEUCYSLEUGLNLEU1A1A1A1ASERTIPVALTHRG1YPHE 138

419 ACCCTCGTTCCTGCCACCTGCTGGCTGTGGCTTCCTTGGGTAACTGGTCTT 460

154 SerasnsrVal,TrpleuserrhrleuthrleugInleuProleucysas 1/

469 ACCAACATCAGCATTCAATCTCTCTTCAACCTCTGGGTACCTCTGTGG 51

170 pproyftva11leasprh1s1rheleucysgluval1pfoalaleuleulysl 18

519 ACACCGCCAGTAGATCACCTTTCTCTGAGATCCAGCACCTCTGCGAT 50

```

18/ eusercysvalglutlthfaldashglualagluLeuphneleuvalser 20
|||||:::|||||:::|||||

```

569 TATCGTGTGATAACCCATGTCAATGAGCTGACCCCTCATGATCACAAGC 61

CC The present sequence is one of a number of isolated polynucleotides  
CC which encode polypeptides involved in olfactory sensation. The  
CC polynucleotides can be used in screening for olfactory agonists and  
CC antagonists. The methods allow for the determination of primary  
CC scents and the identification of the odour receptors used to detect  
CC these primary scents. The methods also enable determination of

601 TTTGTCAGTGCATTATTCTTCTTGTAACCTGTTGCATTAAATCATATTCTC 65

```
seq_documentation_block:
ID    AAH32306 standard; DNA; 1011 BP.
```

CC and edit scents. Libraries of olfactory receptors are useful for  
CC determining the interaction pattern of a composition with the receptors  
CC and can be used for determining differences in the olfactory faculties  
CC of different individuals.

XX  
Sequence 1011 BP; 217 A; 252 C; 202 G; 340 T; 0 other;

alignment\_scores:  
Quality: 918.00 Length: 301  
Ratio: 3.600 Gaps: 0  
Percent Similarity: 84.718 Percent Identity: 58.804

alignment\_block:  
US-09-755-017-2 x AAH32306 ..

Align seg 1/1 to: AAH32306 from: 1 to: 1011

```
5 AaNaPserIleIleGlnIuPheIleuLeuGlyPheSerAspArgp 21
||||| : : : ||||||| : : : : : |||
10 AATGAGAGTTTTGAGAGACTTTTATCTTACTGTGGATTTTCTAACGGCC 59
21 oTrpLeuGluPheProLeuValValPheLeuIleSerTyrrhValt 38
| ||||| ||||| : : : ||||| ||| : : : |||
60 TCATCTGGAGTAGTCTCTCTTGTGGTTACTTGTGACTTCTACTATATA 109
38 hPLePheGlyAsnLeuThrIleIleLeuValSerArgLeuAspThrIys 54
||||| : : : ||||| : : : : : |||
110 CACTGATAGAGAAACCTGTTCATCATCATCTGTCATACCTGACCTCCAT 159
55 LeuHisThrPrometylrPhePheLeuThrAsnLeuSerLeuAspIle 71
||||| : : : ||||| : : : : : |||
160 CTCACACACCCCATGACTTCTTCCTTCAATCTCTCATTTCTCGATGT 209
71 uCysTyrrThrTrhCysThrValProGlnMetLeuValAsnLeuCysSerI 88
||||| : : : ||||| : : : : : |||
210 CTGCTACACACCACGCTCATCCCTGACTGTCTGGTGAATCTCTGGGCC 259
88 lAArgIysValIleSerTyrrArgGlyCysValAlaGlnPheIlePhe 104
: : : : : ||||| : : : : : |||
260 CGGAAGAGACCATCTCTTATGGCGGTGTGACAGTTCACTTACTTGTGTT 309
105 LeuAlaLeuGlyAlaThrGluTyrrLeuLeuAlaValMetSerPheAs 121
||||| : : : ||||| : : : : : |||
310 CTCGCACTGGGAACCCGACAGTGTCTCTCAATCTGTTGATGTCTATGA 359
121 pArgPheValAlaIleCysArgProLeuHisTyrrSerValIleMetHis 138
||||| : : : ||||| : : : : : |||
360 TCGTATAGGAGCTGTGTGAGCTTGTGACATTTCACATGCTCATATCACC 409
138 lAArgLeuCysLeuGlnLeuAlaAlaAlaIleSerTyrrAlaThrGlyPheSer 154
||||| : : : ||||| : : : : : |||
410 CTCGGTTTCGCGCGCTTGTGGCGGGCTTCTGGGTAGTGGTTTACAC 459
155 AsnSerValTyrrLeuSerThrLeuThrLeuGlnLeuProLeuCysAsp 171
: : : : : ||||| : : : : : |||
460 ACCTCAGCACTTCATTCCTCTTACTTCTGATACCCCTATGATGACACA 509
171 oTyrrValIleAspHisPheLeuCysGlnValProAlaIleLeuLysLeu 188
: : : : : ||||| : : : : : |||
510 TCGCCATAGGAGATCACTTCTTCTGTGAGAGCTCAGACATTTCTGCGATT 559
188 eTcysValAluThrThrAlaAsnGluAlaGluPhePheLeuValIleSerGlu 204
||||| : : : ||||| : : : : : |||
560 CAGGTGTGTGATACCTTAGGCAATAGAGACTGACCTCATGGTCAIAGACTGC 609
205 LeuPheHisLeuIleProLeuThrLeuIleLeuIleSerTyrrAlaPheI 221
: : : : : ||||| : : : : : |||
610 ATTTTGTGTTCTATACCTCTCATCTCATCTCATCTCTCATGTTGCCAT 659
221 eValAlaGluAlaValLeuArgIleGlnIleSerAlaGluGlyArgGlnLysAla 238
||||| : : : ||||| : : : : : |||
660 TGGCCGGGGCTGTACTGAGCATGCATCAACCACTGTGGGCTTCAAAAGATGG 709
```

Seq	Sequence	936 BP	199 A	236 C	193 G	308 T	0 other
alignment_scores:	Quality: 912.00						
	Ratio: 3.707						
Percent Similarity:	83.673						
Percent Identity:	60.204						
alignment_block:							
US-09-755-017-2	x AAH32314						
Align seg 1/1	to: AAH32314	from: 1	to: 936				
12	PheilleleuLeuglyPheSerAspArgProTripleuGluPheProleuLe	28					
37	TTTTATTCTACTGGATTTCTTAATTTGGCCCTCAGCTAGACTAGTTCTT	86					
28	uValValPheuileuSerTyrrThrValThlePheGlyAsneuThri	45					
87	TGTGGTTATCTTGATCTTCTTACTCGATGACACTGACAGAACTGTTC	138					
45	lelleleValSerArgglaSphThrylsLeuHistrPrometTyrrPhe	61					
137	TCATCATTCCTCTACGTAGTGACCTCCATCTCCACACACCAATGTACTTC	166					
62	PheleuthrAsneuSerleuLeuAspleucysTyrrThrCysThra	78					
187	TTCCCTTTCACACCTCTCATTTCTTGATCTCGACACACACACAGCTCTAT	236					
78	lProGlmethleuValAsneuLeucysSerlleArglyVallleSerTyra	95					
237	CCCTCAGTGTGGTGAATCTCCGGGCCCGGGAAGAACATCTCGTATG	286					
95	rglyCysValAlaGluPheThlePheleuAlaGlyAlaThrglu	111					
287	CTGTGTGATGTCCTTCACTTACTTGTCTTCTTGACACTGGAAATCGCAG	336					
112	TyrleuLeuLeuAlaValMetSerPheAspArgPheValAlaIleCysar	128					
337	TGTGTCTACTGCTGGTGGATGTCGTATGATGATGATGATGATGATG	366					
128	gProleuHistrGierVallleMetHlsGlnAlaGlyLeuGluLeu	145					
387	ACCTTTCGATTACACTGTCCACAGCACCCCTGCTTTCGCACTGTGG	436					
145	laAlaAlaSerTriValThrclyPheSerAsnSerValTriPleuSerThr	161					
437	CTGGGGCTCTTGGGGTAAATGGTTTACTATCTCAGCACTTCAATCTCC	486					
162	LeuthrleuGluLeuProleuCysAspProTyrrVallleAspHisPhele	178					
487	TTTACTTTCTGGGACCCCTTTGGTGGACATCGCCTATGATGATCACTCTT	536					
178	uGysGlyValProAlaLeuLeuLysLeuSerCysValGluThrThAla	195					
537	CTGGAAGTTCACACACTCTGCGCTTATCATGATGTTGACACCCATGCA	586					
195	snGluAlaGluLeuPheValSerGluLeuPheHisLeuLeuProleu	211					
587	ATGACGACCCCTCATGTGATGATGAGCTCCATTTTGTCTCATACCTTC	636					
212	ThrleuileuLeuSerTyrrAlaPheAlleValAlaValleuArgll	228					
637	ATTCTGATTCACATGCTTGTGGTGCATTTGCCCGGGCTTACTGAGCAT	686					
228	eGlnSerAlaGluGlyArgGlnLysAlaPheGlyThrCysGlySerHisl	245					
687	GCAATCAACACACTGGCTTCAGAAAGTGTAGACATGTCGAGCCCATC	736					
245	euileuValSerleuPheTyrrSerThrAlaValSerValTyrrGlu	261					
737	TTATGCTGTATATCTCTTTTTCATTCACATGTCAGTATATTCACAG	786					

```

seq_name: /SIDS2/gcdata/geneseq/geneseq/NA2001.DAT:AAH32238
seq_documentation_block:
ID   AAH32238 standard; DNA; 1002 BP.
XX
XX   AAH32238;
XX
XX   30-JUL-2001 (first entry)
XX
XX   Human olfactory receptor polynucleotide, SEQ ID NO: 811.
XX
XX   Human; olfactory receptor; OR; primary scent determination;
XX   secondary scent determination; polypeptide library; odour receptor;
XX   scent profile; scent fingerprint; scent representation; ds.
XX
XX   Homo sapiens.
XX
XX   WO200127158-A2.
XX
XX   19-APR-2001.
XX
XX   06-OCT-2000; 2000MO-US27582.
XX
XX   08-OCT-1999; 99US-0158615.
XX   24-FEB-2000; 2000US-0184809.
XX
XX   (DIGI-) DIGISCENTS.
XX   (YEDA ) YEDA RES & DEV CO LTD.
XX
XX   Bellenson J, Smith D, Iancet D, Glusman G, Fuchs T, Yanai I;
XX   WPI: 2001-290713/30.
XX
XX   New polynucleotides which encode polypeptides involved in olfactory
XX   sensation for identifying olfactory agonists and antagonists -
XX
XX   Claim 8; Page 504; 1857pp; English.
XX
XX   The present sequence is one of a number of isolated polynucleotides
XX   which encode polypeptides involved in olfactory sensation. The
XX   polynucleotides can be used in screening for olfactory agonists and
XX   antagonists. The methods allow for the determination of primary
XX   scents and the identification of the odour receptors used to detect
XX   these primary scents. The methods also enable determination of
XX   secondary scents and the identification of combinations of odour
XX   receptors that are involved in detecting such secondary scents.
XX   This enables the construction of a scent representation (also called
XX   a scent fingerprint or scent profile), which may be used to re-create
XX   and edit scents. Libraries of olfactory receptors are useful for
XX   determining the interaction pattern of a composition with the receptors
XX   and can be used for determining differences in the olfactory faculties
XX   of different individuals.
XX
XX   Sequence 1002 BP; 225 A; 256 C; 223 G; 298 T; 0 other;
XX
alignment_scores:
Quality: 908.00 Length: 303
Ratio: 3.575 Gaps: 0
Percent Similarity: 83.828 Percent Identity: 56.436

```

```

alignment block:
US-09-755-017-2 x AAH32238 ..

Align seg 1/1 to: AAH32238 from: 1 to: 1002

5 AsnAspserIleIleGIngluPheIleLeuGluGlyPheSerAspArgp 21
||| ||| : : : : : ||||| : : : : : |||||
19 AAATACAAGTCTTTGCGAAGGCTTCATCTCGTGGGGCTTCTGCATGCTCC 68
||| : : : : : ||||| : : : : : |||||
21 oTrPLeuGluPheProLeuLeuValAlaPheLeuIleSerTyrrhValr 38
| ||||| : : : : : ||||| ||| ||| : : :
69 CCACTAGACAGCTGATCGCTTGTGGTGGTCTCATCTTTATTCCTGTA 118
|| : : : : : ||||| : : : : : |||||
119 CTCTTCTGGCAACATGACCAATGCTTGCTTGGCTTCAAGCTCGGATGCCGG 168
|| : : : : : ||||| : : : : : |||||
55 LeuHisThrProMetTyrrPhePheLeuThrAsnLeuSerLeuLeuAspLe 71
||| ||||| : : : : : ||||| : : : : : |||||
169 CTCACACACCAACCAATGTAATTTCTTTTGGCAACCTCTCAATTCCTGACAT 218
||| : : : : : ||||| : : : : : |||||
71 uCysTyrrhThrCysThrValProGlnMetLeuValAlaLeuCysSerI 88
||| : : : : : ||||| : : : : : |||||
219 GTCTTTCACACACAGGTTCCATCCCTCAAGATGCTCTACACCTTTGGGGCTC 268
||| : : : : : ||||| : : : : : |||||
88 LeuArglyValIleSerTyrrArgglyCysValAlaGlnPheIlePhe 104
||| : : : : : ||||| : : : : : |||||
269 CAGATTAAGACCATATGCTATGTGGTGTGGCATCCAGCTGATCTTGTTC 318
||| : : : : : ||||| : : : : : |||||
105 LeuAlaLeuGlyAlaThrGluTyrrLeuLeuLeuAlaValMetSerPheAs 121
||| : : : : : ||||| : : : : : |||||
319 CTGGCCCTGGGAGGGGGTGGAGCTGTCTCTCGGCTGTCAATGGCAATATGA 368
||| : : : : : ||||| : : : : : |||||
121 ParGpHeValAlaIleCysArgProLeuHisTyrrSerValIleMetHisG 138
||| : : : : : ||||| : : : : : |||||
369 CCGCTAATGCTGCAGCTCTGGCAAAACCCCTGACATACACCAATCATCATCCACC 418
||| : : : : : ||||| : : : : : |||||
138 LInrPLeuCysLeuGlnLeuAlaAlaIleSerTrpValrhnGlyPheSer 154
||| : : : : : ||||| : : : : : |||||
419 CAGCTCTCTGTGGACAGCTGGCTTTCAGTGGCATGGCTGATGGCTTTGGC 468
||| : : : : : ||||| : : : : : |||||
155 AsnSerValTyrrLeuSerThrLeuThrLeuGlnLeuProLeuCysAsprr 171
||| : : : : : ||||| : : : : : |||||
469 AATCTCTCTAATATGGACACCCAGACCAATGAATGATACCCCGGTGGGGCA 518
||| : : : : : ||||| : : : : : |||||
171 oTyrrValIleAspHisPheLeuCysGluValProAlaLeuLeuLysLeus 188
||| : : : : : ||||| : : : : : |||||
519 CAGACGAGTGGACCACTTCTCTGTGAATGACACCACTAAATTTGGATGG 568
||| : : : : : ||||| : : : : : |||||
188 erCysValIGluThrThrAlaAsnGluAlaGluLeuPheLeuValSerGlu 204
||| : : : : : ||||| : : : : : |||||
569 CCGTGTGTAGACACCATGATGCTTAGGAGCATGGCTTTTCCCTGGCAATC 618
||| : : : : : ||||| : : : : : |||||
205 LeuPheHisLeuIleProLeuThrLeuIleLeuIleSerTyrrAlaIle 221
||| : : : : : ||||| : : : : : |||||
619 TTTATATCATCTGGGACACCATCATCTCATCTTCATTTCTTAAAGGTAACT 668
||| : : : : : ||||| : : : : : |||||
221 eValArgAlaValLeuArgIleGlnSerAlaGlnValArgGlnLysAlaP 238
||| : : : : : ||||| : : : : : |||||
669 TGGAGGACAGTGTGTAGGATCAACTGACTGCTGGGCGAAAGAAAGGCTT 718
||| : : : : : ||||| : : : : : |||||
238 heGlyThrCysGlySerHisLeuIleValSerLeuPheTyrrSerThr 254
||| : : : : : ||||| : : : : : |||||
719 TCACACACTTGCAGCTCGCATCTAATTTGTGTCTCTCTTCTTATGGTACA 768
||| : : : : : ||||| : : : : : |||||
255 AlaValSerValTyrrLeuGlnProProSerProSerSerLysAspGln 271
||| : : : : : ||||| : : : : : |||||
769 ATCATATATACATGTAATCTCCACACCAATATCTTATTTCCAGAGACAGGG 818
||| : : : : : ||||| : : : : : |||||
271 yLysMetValSerLeuPheTyrrGlyIleIleAlaProMetLeuAsnProL 288
||| : : : : : ||||| : : : : : |||||
819 CAAAGTTCTTACCCCTTTTCTTCAACAAATGTACTGCCCACTGTATTAACCCC 868
||| : : : : : ||||| : : : : : |||||

```

288 euileTyrrhrLeuArgAsnLysGluValLysGluGlyPheLysArgLeu 304  
|||||  
869 TGATCTATACACTAAGAAACAAGATGTAAAGAGGCCCATGAAGAAGTGC 918  
305 ValAlaArg 307  
919 CTAGGGAAG 927



OM of: US-09-755-017-2 to: EST:\*. out\_format: pfs  
Date: Feb 26, 2002 5:25 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

-MODEL=framed.p2n.model -DEV=xlh  
-Q=/c9g2\_1/USPTO.spool/US09755017/runal\_25022002.165640.20760/app\_query.fasta\_1.374  
-DB=EST -DBMT=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000  
-MIMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500  
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=0.000 -YGAPEXT=10.000 -YGAPEXT=0.500 -DELDP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blomsun62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500  
-MILEN=0 -MILEN=2000000000 -USER=US09755017.ecgn1.1.3653  
-NCPU=6 -ICPU=3 -LONGLOGS -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPHY -WAIT -THREADS=1

## Search information block:

Query: US-09-755-017-2  
Query length: 313  
Database: EST:  
Database sequences: 11351937  
Database length: 1077921985  
Search time (sec): 1205.400000

## score\_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_hlc:AK016560	+	605.50	933.98	2.7e-42	3063	AK016560 Mus musculus adult ma
gb_hlc:AK017036	+	604.50	930.01	1.2e-42	1394	AK017036 CIT-HSP-2354D1.TF CIT-
gb_gss:AK0077154	-	573.00	891.18	1.8e-40	479	AK0077154 CIT-HSP-2354D1.TF CIT-
gb_gss:AK0090606	-	567.50	879.04	8.5e-40	692	AK0090606 RPT-23-27B20.TV RPT-
gb_gss:BH111304	+	566.50	876.23	1.2e-39	788	BH111304 RPT-24-16786.TV RPT-
gb_gss:AK0096227	+	554.00	859.73	1.1e-38	642	AK0096227 2M024J24R.Mouse 10kb
gb_gss:AK0255734	+	544.00	841.91	9.9e-38	726	AK0255734 RPT-23-16619.TV RPT-
gb_hlc:AK016338	+	534.00	834.90	2.4e-37	1501	AK016338 Mus musculus adult me
gb_est1:AI604336	-	534.00	829.48	2.4e-37	523	AI604336 vv74c09.xl Stratagene
gb_gss:AK2913406	-	509.50	787.22	1.1e-34	794	AK2913406 RPT-23-165C12.TV RPT-
gb_gss:AK2535949	-	505.50	782.53	2.0e-34	676	AK2535949 1M009J01F.Mouse 10kb
gb_gss:AK2909618	-	497.50	766.46	1.2e-33	797	AK2909618 RPT-24-222618.TV RPT-
gb_est2:BG193339	+	493.00	761.87	3.8e-33	762	BG193339 RST12467 Athersys RAGE
gb_gss:AK2607393	+	492.00	760.55	3.3e-33	740	AK2607393 1M042N15R.Mouse 10kb
gb_est2:BG197640	+	490.50	757.55	4.9e-33	796	BG197640 RST17016 Athersys RAGE
gb_gss:AK2664867	+	479.00	743.46	3.0e-32	534	AK2664867 1M054E05R.Mouse 10kb
gb_gss:AK2511623	-	475.00	735.59	8.3e-32	632	AK2511623 1M035C17R.Mouse 10kb
gb_gss:AK2086625	-	470.50	727.78	2.3e-31	686	AK2086625 RPT-23-26E4.TV RPT-2
gb_gss:AK2974496	+	468.50	724.53	3.4e-31	695	AK2974496 2M024B07F.Mouse 10kb
gb_gss:AK2711125	+	467.00	724.49	3.4e-31	548	AK2711125 RPT-23-14624.TV RPT-
gb_gss:AK0482125	+	466.50	723.62	3.8e-31	553	AK0482125 RPT-11-261E3.TV RPT-
gb_gss:AK2638594	+	465.00	719.78	6.3e-31	646	AK2638594 1M056E13R.Mouse 10kb
gb_gss:AK2964406	+	463.00	716.19	1.0e-30	678	AK2964406 2M023A00R.Mouse 10kb
gb_gss:AK2114477	+	460.00	711.45	1.8e-30	682	AK2114477 RPT-23-9A5.TV RPT-23
gb_gss:AK2914130	-	457.00	705.95	3.7e-30	743	AK2914130 RPT-24-164M20.TV RPT-
gb_gss:AK2103967	-	456.00	704.71	3.9e-30	656	BH078380 RPT-24-349E9.TV RPT-
gb_gss:AK2720443	-	455.00	704.71	4.3e-30	611	AK2720443 RPT-23-33G18.TV RPT-
gb_gss:AK2765752	+	454.00	702.83	5.5e-30	632	AK2765752 1M056E13R.Mouse 10kb
gb_gss:AK2109687	+	453.50	702.42	5.8e-30	608	AK2109687 RPT-23-13C2.TV RPT-2
gb_gss:AK207832	+	448.00	693.67	1.8e-29	619	AK207832 1M017B09R.Mouse 10kb
gb_gss:AK252656	+	446.00	688.67	3.4e-29	752	AK252656 RPT-23-210N3.TV RPT-
gb_gss:AK2019257	-	445.50	684.51	3.0e-29	629	AK2019257 RPT-23-26H21.TV RPT-
gb_est1:AV693866	-	442.00	683.73	5.0e-29	534	AK231830 1M002J19R.Mouse 10kb
gb_gss:AK2312830	+	436.00	675.10	1.9e-28	609	AK2312830 RPT-23-231B2.TV RPT-

gb\_gss:BH069789 - 436.00 673.78 2.3e-28 698 | BH069789 RPT-24-331A10.TVB  
gb\_gss:AK2648626 + 435.00 677.18 2.5e-28 632 | AK2648626 1M0517B18R.Mouse 1  
gb\_gss:AK2725747 - 434.50 663.63 3.9e-28 842 | AK2725747 RPT-24-98112.TVB  
gb\_gss:BH078443 - 433.50 670.73 3.4e-28 639 | BH078443 RPT-24-349E9.TV R  
gb\_est1:AV689607 - 433.50 670.13 3.7e-28 680 | AV689607 AV689607 GKC Homo

seq\_name: gb\_hlc:AK016560

seq\_documentation\_block:

LOCUS AK016560 3063 bp mRNA HTC 05-JUL-2001  
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4932441H21, full insert sequence.

ACCESSION AK016560  
VERSION AK016560.1 GI:12855357  
KEYWORDS CAP trapper.

Mus musculus (strain:C57Bl/6J) adult male testis cDNA to mRNA, clone:4932441H21.  
clone:4932441H21.

## ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
1 (bases 1 to 3063)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Methods in enzymology. 303, 19-44 (1999)  
99279253

MEDLINE  
10349636  
2 (bases 1 to 3063)

REFERENCE  
2 (bases 1 to 3063)  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome research. 10 (10), 1617-1630 (2000)  
20499374  
11042159

MEDLINE  
11042159  
3 (bases 1 to 3063)

REFERENCE  
3 (bases 1 to 3063)  
Shibata, K., Itoh, M., Aizawa, K., Nagoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunaka, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsuno, H., Sakaguchi, S., Ikegami, T., Kaishiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneeda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multiplexed sequencer  
Genome research. 10 (11), 1757-1771 (2000)  
11076861

MEDLINE  
20530913  
4 (bases 1 to 3063)

REFERENCE  
4 (bases 1 to 3063)  
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
5 (bases 1 to 3063)

MEDLINE  
11076861  
5 (bases 1 to 3063)

REFERENCE  
5 (bases 1 to 3063)  
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arikawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagata, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Icomori, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shingawa, A., Shitaki, T., Soabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toyai, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (10-JUL-2000) Yoshida Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp, URL: http://genome-gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

TITLE  
JOURNAL  
Submitted (10-JUL-2000) Yoshida Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp, URL: http://genome-gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)



ORGANISM	Mus musculus Clone:4933433E02.
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 1394)
AUTHORS	Carninci,P. and Hayashizaki,Y.
JOURNAL	High-efficiency full-length cDNA cloning Methods in enzymology. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	2 (bases 1 to 1394) Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome research. 10 (10), 1617-1630 (2000)
TITLE	
JOURNAL	
MEDLINE	20499374
PUBMED	11042159
REFERENCE	3 (bases 1 to 1394) Shibata,K., Itoh,M., Aizawa,K., Nagoka,S., Sasaki,N., Carninci,P., Konno,H., Ariyama,J., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A., Yamanoto,R., Matsunoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yonezaki,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome research. 10 (11), 1757-1771 (2000)
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	20530913
PUBMED	11076861
REFERENCE	4 (bases 1 to 1394) The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001) 5 (bases 1 to 1394)
TITLE	
JOURNAL	
MEDLINE	
PUBMED	
REFERENCE	Aachi,J., Aizawa,K., Akaitira,S., Akimura,T., Aono,H., Arai,A., Atakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiroaka,T., Horii,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ono,M., Okazaki,Y., Orkido,T., Owa,C., Saio,H., Saio,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,K., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,T., Toyota,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y. Direct Submission Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gscc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
PUBMED	
REFERENCE	Please visit our web site ( <a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a> ) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GGAAGAGAAGATCCAGACCTCTTTTGTATTTCVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot -10.0 and subtraction to Rot =100.0. Second strand cDNA was prepared with the primer adaptor of sequence [5' GGAGAGCAATTCCTCACTTAATTAAATTAATCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size comprised between 0.5 and 3 kb was selected before cloning. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI;
COMMENT	

```

FEATURES                      3 end: BamHI, Host: DH10B,
                               Location/Qualifiers
                               1..1394
                               /organism="Mus musculus"
                               /strain="C57BL/6J"
                               /db_xref="taxon:10090"
                               /db_xref="MGD:MGI:1894709"
                               /db_xref="MGD:MGI:1914036"
                               /clone="493343E02"
                               /sex="male"
                               /tissue-type="testis"
                               /clone_lib="RIKEN full-length enriched mouse cDNA library"
                               /dev_strag="adult"
                               332..1282
                               /note="putative"
                               /codon_start=1
                               /protein_id="BAB30564.1"
                               /db_xref="GI:12856092"
                               /translation="MEPNNMDESGFLLNGTLOGSSPELLCAVNTLYMALISNGIE
                               LLVYTDARLHVPMYLLRLSLIDILFTSVYTPRVNVDLNDNTISFGCALQLE
                               SAMTLGGAEDELLEFMAVDRYVALCHPLNVTIEMSPKRLMAIIMILASLNGHT
                               VYTMHFFPESQDEIRLHCEVPLRLACADTSQELMVTYGVIFLLPLSAITIS
                               SLIFETVLHMPSENGRKKAIVTCSHSHTLVGMFEFGATFMVLPSSPSPKODNIISV
                               FYETVPLALNDPLYSIENKKEVIGAVRVLGHHILPAHATV"

BASE COUNT      355 a      337 c      304 g      398 t
ORIGIN

alignment_scores:
    Quality: 604.50      Length: 308
    Ratio: 2.663      Gaps: 1
    Percent Similarity: 73.701      Percent Identity: 40.260

Alignment block:
US-09-755-017-2 x AK017036 ..

Align seg 1/1 to: AK017036 from: 1 to: 1394

3 TTPVALASNPSPSERILEELGNGLPHLEULEUAGLYPHESERAS 19
||| |||:||||: ||: |||:||||:||||: ||
341 TGG...AATTCTAACCTTGGAAGATGGATCATCTTGCGGATCTGGA 387
19 rATGPrCTrPLeuGluPhePProLeuValAlaPheLeuIleSerTyr 36
| ||| |||: |||: |||: |||: |||: |||: |||
388 TGGCAGTGGCTCTCTGGAAGCTGCTGTGCACAGTTACACCCCTGTAA 437
36 hValThrIlePheGlyAsnLeuThrIleIleLeuValSerArgLeuAsp 52
||||:||||:||||:||||:||||:||||: |||:||||:||||: |||
438 TGTGGCAGTGCATCAGCAATGCACTTCTACTCTCTGCTGCATCAGAGTGAT 487
53 ThrIleLeuIleThrProMetTyrPhePheLeuThrAsnLeuSerLeuLe 69
||||:||||:||||:||||:||||:||||: |||:||||:||||: |||
488 GGCGGGCTTCAGATCCCATGTAACCTCTGTAAGGAGAGCTGTCTGCAT 537
69 uAspLeuCysTyrThrThrCysThrValProGlnMetLeuValAsnLeu 86
||||| |||:||||: |||:||||: |||:||||: |||:||||: |||
538 TGAACCCCTCTCACAATCAGTTGTAACCTCCAAACAGTGTGTGGATTTTC 587
86 ySerSerLeuArgValIleSerTyrArgGlyCysValIleGlnLeuPhe 102
||||:||||:||||:||||:||||:||||: |||:||||:||||: |||
588 TGTCTAGAGACACACCAATATCTTTGAGGAGATGTCCTCTCAATGTTTC 637
103 IlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuLeuAlaValMetSe 119
||||:||||:||||:||||:||||:||||: |||:||||:||||: |||
638 TCAGCAATGACATTTGGGTGGTGAGAGAGACCTCTCTGGCCCTTCATGGC 687
119 rPheAspArgPheValAlaIleCysArgPProLeuIleTyrSerValIleM 136
||||:||||:||||:||||:||||:||||: |||:||||:||||: |||
688 CTATGATAGATATGTGGCCATTTGCATCTCTCAAACTACATGATCTCTTA 737
136 eHisIleGlnArgLeuCysLeuGlnIleLeuAlaAlaIleSerThrValIle 152
||| |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
738 TGAAGTCCAAAGGCTCTCAGGCTCATGTGTGGCCATATCATGATGATCTCA 787

```

[nccepr://www.ncbi.nlm.nih.gov/nucleotide/bac\\_end\\_seq.cif/bac\\_end\\_seq.cif](#).  
Seq primer: M13-21

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;



/clone\_lib="RPCI-24"  
 /sex="Male"  
 /cell\_type="Spleen/Brain"  
 /note="Vector: pTARBAcl; Site\_1: BamHI; Site\_2: BamHI;  
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The  
 library was cloned in the pTARBAcl cloning vector at the  
 BamHI sites using MboI partially digested male C57BL/6J  
 DNA."  
 BASE COUNT 168 a 217 c 162 g 241 t  
 ORIGIN

alignment\_scores:  
 Quality: 566.50 Length: 265  
 Ratio: 2.737 Gaps: 7  
 Percent Similarity: 78.113 Percent Identity: 46.415

alignment\_block:  
 US-09-755-017-2 x BH11304 ..

Align seg 1/1 to: BH11304 from: 1 to: 788

```

35  TTTTThValThrllePhelglsnleuThrlleleuValSerArgle 51
    ||| : : : : : ||| : : : : : |||
6  TACATCTTCAGCCTCGCAATGGAATAGATAGTATGCTATGCTTCC 55
    ||| : : : : : ||| : : : : : |||
51  uasprThllysluHsthrProketyrPhelauThrAsnleuSerl 68
    ||| : : : : : ||| : : : : : |||
56  GGAACCCCAAGCTGCACACCCCTATGACTTCTTCTTCATCTG6CCA 105
    ||| : : : : : ||| : : : : : |||
106 TCTTGACATATCTATGCTTCCACAAATGTCGCCAAGATGTTGCCAAC 155
    ||| : : : : : ||| : : : : : |||
85  LeuGysSerlleArglyValIleSerTyraGlyCysValAlaIle 101
    ||| : : : : : ||| : : : : : |||
156 CTTATTAACAGCAAGAAAGAACATCTATGCCCCCATGCATCACACAGAC 205
    ||| : : : : : ||| : : : : : |||
101 upheillePhelAuAlaLeuGlyValThrglyluThleuLeuAlaValM 118
    ||| : : : : : ||| : : : : : |||
206 ATTCTTGATTTGGCTTGTGCGCTTGGAGTCTTGATTTTGGCAGCCA 255
    ||| : : : : : ||| : : : : : |||
118 eIseRheAspArgPheValAlaIleCysArgProleuHstIySerIy 134
    ||| : : : : : ||| : : : : : |||
256 TGGCCATGACAGAGTTGTGGCCATCTGCGACCCCTACATCACTACT 305
    ||| : : : : : ||| : : : : : |||
134 llekethIsglnArgLeuGlyCysleuGlnleuAlaAlaIleSerTy 151
    ||| : : : : : ||| : : : : : |||
306 CATCATGAGTTGGAAAGTATGTGTGCTGCTGCTGCATCTTGGTCAT 355
    ||| : : : : : ||| : : : : : |||
151 hGlyPheSerAsnSerValITrPLeuSerThrlleuThrlleuGlnleuPro 167
    ||| : : : : : ||| : : : : : |||
356 GTGGATTTAGCCTCTCTGTGGACACACATTTCTCTGTAAGGTTGCC 405
    ||| : : : : : ||| : : : : : |||
168 LeuGysAspProTyraIleAspHisPheLeuGlyValProAlaIle 184
    ||| : : : : : ||| : : : : : |||
406 TTTTGGGGCCCCAGGAATTAACACCTTCTGTGAATCTCGCTGT 455
    ||| : : : : : ||| : : : : : |||
184 uLeuLySerleuSerCysValIgluThrAlaAsnGlnAlaIleuPhe 201
    ||| : : : : : ||| : : : : : |||
456 TCTCAAGCTGGCCCTGCGCACACTTGTATCAACCAATC...GTCA 502
    ||| : : : : : ||| : : : : : |||
201 euValSerleuPheHisleuIle...ProleuThrlleuIleleu 216
    ||| : : : : : ||| : : : : : |||
503 TACGTCAAGTGTCTTGTCTTGTAGTGAACCCCTTGTCACTGTGT 552
    ||| : : : : : ||| : : : : : |||
216 eSerTyraIlePheIleValAlaIleValAlaIleuArgIleGlnleu 233
    ||| : : : : : ||| : : : : : |||
553 ATCTTACAGCACATCTCTGACATCTCTGAATAATGCTCAAGAGG 602
    ||| : : : : : ||| : : : : : |||
233 lYArgIlnAlaIlePheGlyThrCysGlySerHisleuIleValSer 249
    ||| : : : : : ||| : : : : : |||
603 GTGCGAGGAAGGCTTTTCACCTTGTCTCCATCTCTGTGTGGTGG 652
  
```

```

250 LeuPheTySerThrlAlaValSerValTyrlleuGlnProSerPro 265
    ||| : : : : : ||| : : : : : |||
653 CTCTTCTTTGGTATAGCATGCTGGTGTATATAGTCCCTGACTGTGAT 702
    ||| : : : : : ||| : : : : : |||
266 SerSerLyAspGlnGlyMetValSerLeuPheTyrlleuIleAl 282
    ||| : : : : : ||| : : : : : |||
703 CACGAGAGAGAGAGAGAAATTTTGTCTCAGTCCACAGTCTTTTAC 752
    ||| : : : : : ||| : : : : : |||
282 aProMetLeuAsnProleuIleTyrlleuAlaGln 294
    ||| : : : : : ||| : : : : : |||
753 CCAAGTGAATTCCT...CATATAGTCTGAGAGAC 786
    ||| : : : : : ||| : : : : : |||
  
```

seq\_name: gb\_gss:A2969227

seq\_documentation\_block: 642 bp DNA 27-APR-2001  
 LOCUS A2969227  
 DEFINITION 2M0241J24R Mouse 10kb plasmid U0GC2M library Mus musculus genomic  
 clone U0GC2M0241J24 R, DNA sequence.  
 ACCESSION A2969227  
 VERSION A2969227.1 GI:13840454  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 642)  
 Dunn,D., Moyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Kelly,  
 M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A.  
 and Wright,D., Weiss,R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah  
 Department of Biology  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0241 row: J column: 24  
 Seq primer: CACACAGGAACAGCTATGACAC  
 Class: plasmid ends  
 High quality sequence stop: 642.  
 Location/Qualifiers  
 1..642  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="U0GC2M0241J24"  
 /clone\_lib="Mouse 10kb plasmid U0GC2M library"  
 /sex="Female"  
 /lab\_host="E. coli strain XL10-Gold, Tl-resistant, F-"  
 /note="Vector: pMD42rv; Purified genomic DNA from M.  
 musculus C57BL/6J (female) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pMD42 (911473211419b1AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into

#### FEATURES

source

chemically-competent *E. coli* XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 130 a 166 c 129 g 217 t  
ORIGIN

alignment\_scores: Length: 213  
Quality: 554.00 Gaps: 0  
Ratio: 3.240  
Percent Similarity: 80.282 Percent Identity: 47.418

alignment\_block:  
US-09-755-017-2 x AZ969227 ..

Align seg 1/1 to: AZ969227 from: 1 to: 642

```

79 ProGlnMetLeuValAsnLeuCysSerIleArgLysValIleSerTyrAr 95
   |||||.....|||.....|
3  CCCAGATGATGTCATCTCTGTCAGAGAAATCAATTCCTATGG 52
95 gGlyCysValAlaGlnLeuPheIlePheLeuAlaLeuGlyAlaThrGluT 112
   |||||.....|||.....|
53 AGGCTGCTGACCAAGCTCTTGCATTTCTTGTGGCTCAGAGT 102
112 yrlLeuLeuAlaValMetSerPheAspArgPheValAlaIleCysArg 128
   |||||.....|||.....|
103 GTCCTCTCTGGCAGCATGTCATATGTCATATGTCATCTCTAG 152
129 ProLeuHisTyrSerValIleMetHisGlnArgLeuCysLeuGlnLeuAl 145
   |||||.....|||.....|
153 CCGTTAAGGTAAGTATATATGTAACAAGCTCTGTCAGCTGCTAGC 202
145 AlaAlaSerIleValThrGlyPheSerAsnSerValIlePheLeuSerThr 162
   |||||.....|||.....|
203 AGCTTCATGCTGACAGCTGGGTTTCTCACTGATGTCACACAGGTTT 252
162 eurtHleuGlnLeuProLeuCysAspProTyrValIleAspHisPheLeu 178
   |||||.....|||.....|
253 TGACCTTCACACTGGCCCTTTGTGTGTACAAATCAATTAATTTCTTC 302
179 CysGluValProAlaLeuLeuLysLeuSerCysValGluThrThrAlaAs 195
   |||||.....|||.....|
303 TGTGCAATCACTCCCTTGTGTCATCTTGTGTGATACATCTCCCTCAA 352
195 nglAlaGlnLeuPheLeuValSerGlnLeuPheHisLeuLeuProLeuT 212
   |||||.....|||.....|
353 TGAATGCTTGTGCTGTCATTTGGATCCCTCATAGGCTGACCTCTTTC 402
212 hrLeuIleLeuIleSerTyrAlaPheIleValArgAlaValLeuArgIle 228
   |||||.....|||.....|
403 TGTGGTATCTCTTCTTACCTTTTACATCATCTCCACATCTCGAGATC 452
229 GlnSerAlaGlnGlyArgGlnLysAlaPheGlyThrCysGlySerHisIle 245
   |||||.....|||.....|
453 CGTTCCTCTGAGGGGAGGACAAAGCTTTTCCACCTGTCGCCACACT 502
245 uileValValSerLeuPheTyrSerThrAlaValSerValTyrLeuGlnP 262
   |||||.....|||.....|
503 GCTCATTTGTTATCTCTATATAGGAGTATCTTCAAGTATGAGAGC 552
262 roProSerProSerSerLysAspGlnGlyLysMetValSerLeuPheTyr 278
   |||||.....|||.....|
553 CCATCTCATCTTACTCTCTAGAGAAAGATAGATTGATCTCAGTCTGAT 602
279 GlyIleIleAlaProMetLeuAsnProLeuIleTyrThr 291
   |||||.....|||.....|
603 AGTGTTCACACCCATGCTGATCTGTAATTTATACG 641

```

seq\_name: gb\_AZ255734

seq\_documentation\_block: 726 bp DNA GSS 26-JUL-2000  
LOCUS AZ255734  
DEFINITION RPCI-23-16619, 'J' RPCI-23 Mus musculus genomic clone RPCI-23-16619,

ACCESSION DNA sequence.  
VERSION AZ255734  
KEYWORDS GI:9458784  
SOURCE GSS.  
ORGANISM Mus musculus  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 726)  
Zhao, S., Nieman, W., Feldblum, T., Malek, J., Shatsman, S., Akinet  
, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.  
and Fraser, C.M.  
Mouse BAC End Sequences from Library RPCI-23  
Unpublished (1999)  
Other GSSs: RPCI-23-16619.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC  
library availability, please contact Pieter de Jong  
(pieter@jeong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.html)  
or from Resea ch Genetics (info@resgen.com). BAC end page:  
http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html  
Plate: 166 row: I column: 9  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
Location/Qualifiers  
1..726  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-16619"  
/clone\_1fb="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site:1:  
EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and EcoRI Methylase. Size  
selected DNA was cloned into the pBAC3.6 vector at the  
EcoRI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 167 a 171 c 118 g 270 t  
ORIGIN

alignment\_scores: Length: 239  
Quality: 544.00 Gaps: 0  
Ratio: 2.973  
Percent Similarity: 76.569 Percent Identity: 42.259

alignment\_block:  
US-09-755-017-2 x AZ255734 ..

Align seg 1/1 to: AZ255734 from: 1 to: 726

```

26 ProLeuLeuValAlaPheLeuIleSerTyrThrValThrIlePheGlnAs 42
   |||||.....|||.....|
5  CCTCTGTTCTTCTGTTCTGTAATGTAATGTAATGTAATGTAATGTAATG 54
42 nLeuThrIleIleLeuValSerArgPheAspThrThrLysLeuHisThrProm 59
   |||||.....|||.....|
55 TTTGGGTTTGGTAATCTGAATGGAGTGAATGTAATGTAATGTAATGTAAT 104
59 eTyrThrPheLeuThrAsnLeuSerLeuLeuSerLeuLeuSerLysTyrThr 75
   |||||.....|||.....|
105 TGTACTTTTCTCTTAACTGTCCTTGTGATGAGACCTGTGTACTCTTCA 154

```







US-09-755-017-2 x A1604386/rev ..

Align seg 1/1 to reverse of: A1604386 from: 1 to: 523

```

89 ArgLyValIleSerTyrArgGlyCysValAlaGlnLeuPheIlePhe 105
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
517 CGAAAGACATCAGCTTGTGGATGCTTGTGCACGCTTCTTCATCTCTT 468
105 uAlaLeuGlyAlaThrGluTyrLeuLeuLeuAlaValMetSerPhaAsp 122
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
467 ATCTCTGGGACGACGAGTGCATCTGCTAGACAGTGAAGGCTTTGACC 418
122 rrpHValAlaIleCysArgProLeuHisTyrSerValIleMetHisGln 138
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
417 GCTATGTGGTGTGTGCGACGCTGACATGATGCCACATCATCCACCT 368
139 ArgLeuCysLeuGlnLeuAlaAlaSerTyrValThrGlyPheSerAs 155
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
367 CCGCTGTGTGGCCAGCTGGCAGCTGTGGCTGACATTTGGCTAGTGA 318
155 nserValTyrPleuSerThrLeuThrLeuGlnLeuProLeuCysAspPro 172
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
317 GTTAGTAGTTCAGACCCCATCCACTCTGCGCTGCTTGTGCCACC 268
172 yrValIleAspHisPheLeuCysGluValProAlaLeuLeuLysLeuSer 188
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
267 ATCAGGTTCAGCACTTGTGTGTGAGGTCCCTGTTGATTCAGCTGCC 218
189 CysValGluThrThrAlaAsnGluAlaGluLeuPheLeuValSerGlu 205
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
217 TGGGGGACACCCATTAATGAGATCAAAATGCTATTCGACGTCTT 168
205 upHnHisLeuIleProLeuThrLeuLeuLeuLeuSerTyrAlaPheIle 222
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 CATCTGTGTGTACCTCTGAGCTCATCTCTGTCTTATGATGCCATTG 118
222 aAlaGlnAlaValLeuArgIleGlnSerAlaGlnGlyArgGlnLysAlaPhe 238
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 CCAGGCGACAGTGTGAGGATTAAGCTTGCAAAAGGCGCAGAAAGCTTT 68
239 GlyThrCysGlySerHisLeuIleValValSerLeuPheTyrSerThrAl 255
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67 GGGACCTGCTCTCCACCTCATTTGTGTGCTCTTCCACAGCTCAGT 18
255 aValSerValTyrLeu 260
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 CATCGCTGTTATCTG 2

```

seq\_name: gb\_A2913406

seq\_documentation\_block:

LOCUS A2913406 794 bp DNA GSS 05-MAR-2001  
DEFINITION RPCI-24-165C12.TV RPCI-24 Mus musculus genomic clone RPCI-24-165C12  
' DNA sequence.

ACCESSION A2913406  
VERSION A2913406.1 GI:13232351

KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 794)  
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,  
Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,  
Russell, D., de Jong, P. and Fraser, C.M.

TITLE Mouse BAC End Sequences from Library RPCI-24  
JOURNAL Contact: Shaying Zhao  
COMMENT Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208

Email: szhao@ligr.org  
Clones are derived from the mouse BAC library RPCI-24. For BAC  
library availability, please contact Pieter de Jong  
(pdejong@mail.chd.org). Clones may be purchased from BAC PAC  
Resources (<http://www.chori.org/bacpac/orderingframe.html>). BAC end  
page: [http://www.tigr.org/ldb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/ldb/bac_ends/mouse/bac_end_intro.html)  
Plate: 165 row: C column: 12  
Seq primer: 17  
Class: BAC ends.

FEATURES  
source Location/Qualifiers  
1..794

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-24-165C12"  
/clone\_lib="RPCI-24"  
/sex="Male"  
/cell\_type="Spleen/Brain"  
/note="Vector: pPARBAC1; site:1: BamHI; site:2: BamHI;  
RPCI-24 Mouse BAC library produced by Pieter de Jong. The  
library was cloned in the pPARBAC1 cloning vector at the  
BamHI sites using MboI partially digested male C57BL/6J  
DNA."

BASE COUNT 254 a 170 c 221 g 149 t  
ORIGIN

alignment\_scores:  
Quality: 509.50 Length: 259  
Ratio: 2.640 Gaps: 5  
Percent Similarity: 74.517 Percent Identity: 40.154

alignment\_block:

US-09-755-017-2 x A2913406/rev ..

Align seg 1/1 to reverse of: A2913406 from: 1 to: 794

```

27 LeuLeuValAlaPheLeuIleSerTyrThrValThrIlePheGlyAsnLe 43
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
762 CTCTTTCTTTCTCTCTCTCGGTGTATGTGACAACTCTGCTGGAAACT 713
43 uThrIleIleLeuValSerArgLeuAspThrLysLeuHisThr.Promet 59
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
712 TCTCATCATGTCACAGTGCAGCTGTGAGTCCGCTTCCACACCCCATG 663
60 TyrPhePheLeuThrAsnLeuSerLeuAspLeuCysTyrThrTrpCys 76
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
662 TATTTCTGTCTCAGAAATTTGTCTGTGATATGCTTTTCTCCAT 613
76 srHrValProGlnMetLeuValAsnLeuCysSerIleArgValValIleS 93
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
612 CATTCGCCCCANAGCTTCTGTAGACCTCTCTCTGACAGACAAACCATCT 563
93 eTyrArgGlyCysValAlaGlnLeuPheIlePheLeuAlaLeuGlyAla 109
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
562 CTTTCAATGGCTGTCTCAGTCAAGATGTTCTTCCACCTTATGGGGGA 513
110 ThrGluTyrLeuLeuLeuAlaValMetSerPhaAspArgPheValAlaI 126
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
512 GTGGATGTGTTTCTGTGTGATGGCTGTGATGATGATGTGGCCAT 463
126 eCysArgProLeuHisTyrSerValIleMetHisGlnArgLeuCysLeu 143
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
462 CTCTAGCCCTTGCATATGTACACATCAAGACACAGAGGCGTTGCATTG 413
143 ILeuAlaAlaAlaSerTyrValThrGlyPheSerAsnSerValTyrPleu 159
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
412 GGTAAATAGTGGCTCTGCGTGGGGGCTTCCACCTCATTAAGTCAG 363
160 SerThrLeuThrLeuGlnLeuProLeuCysAspProTyrValIleAspH 176
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
362 ATTTCCTCTCTGTCAGACTTCCATTCTGTGACCAATGTTCTGTATAC 313

```

```

176 sPheLeuGysgluValProAlaLeuLeuLysLeuSerCysValgluThrT 193
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
312 TTCTACTGATGATGTCGCCAGCATCACTGCTGCTGACAGACATCT 263
193 hAlaAsnGLuAlaLeuPheLeuValSerGlu.....LeuPheHis 207
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
262 TTGTT.....CTTGAGCTGATGATGATTTCCAAATATAGTGGCTTGCT 219
208 LeuLeuProLeuThrLeuLeuLeuLeuSerTyrAlaPheLeuValArgAl 224
    :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
218 ACTCTGCTGTTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 172
224 aValLeuArgIleGlnSerAlaGluGlyArgGlnLysAlaPheGlyThrC 241
    ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
171 GATGCTTGAATCTCATTTGAGAGAGGCGAGGAAGAACCATCTCCACCT 122
241 ySgLSerHisLeuLeuValValSerLeuPheTyrSerThrAlaValSer 257
    || ||||| ||||| ||||| ||||| ||||| ||||| |||||
121 GCACCTCCACATCACTGTGTGACATGCTGCTGCTGCTGCTGCTGCTAT 72
258 ValTyrLeuGlnProProSerProSerSerLysAspGlnGlyLysMetAla 274
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
71 GCTCTATGCCAGGCGCCCTTCACTGCTCCCTCCCGACGAT.....AAGCCAT 28
274 lSerLeuPheTyrGlyIleIleAla 282
    :|||::: :::: |||||:::
27 CTCTGTACACCTTCACAGTCATCTCC 3
seq_name: gb-ss-A2353949

seq_documentation_block:
LOCUS      A2353949             676 bp      DNA             GSS             02-OCT-2000
DEFINITION IM0093J01F Mouse 10kb plasmid UNGC1M library Mus musculus genomic
ACCESSION  A2353949
VERSION     A2353949.1 GI:10464961
KEYWORDS   GSS.
SOURCE      house mouse.
ORGANISM   Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE   1 (bases 1 to 676)
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
            Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
            Unpublished (2000)
            Contact: Robert B. Weiss
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0093 row: J column: 01
            Seq primer: CGTGTAAACGACGCGCCAGT
            Class: plasmid ends
            High quality sequence stop: 676.
            Location/Qualifiers
                1..676
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UNG1M0093J01"
                /clone_1kb="mouse 10kb plasmid UNGC1M library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                /note="Vector: PMD42nv; Purified genomic DNA from M.
                musculus C57BL/6J (male) was obtained from the Jackson

```

```

Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-replicated with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (9114/732114[9b]AF12907.1), a copy-number
inducible derivative of plasmid RL. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT      173 a      149 c      135 g      219 t
ORIGIN

alignment_scores:
    Quality: 505.50      Length: 232
    Ratio: 2.939      Gaps: 3
    Percent Similarity: 74.138      Percent Identity: 46.552

alignment_block:
US-09-755-017-2 x A2353949 ..

Align seg 1/1 to: A2353949 from: 1 to: 676

74 ThrThrCysThrValProGlnMetLeuValAsnLeuCysSerIleArgly 90
    ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
3 TCCTCAGTACACAGACCACCAATGATGTTGTTGACTGTTACTGTAAGAAAA 52
90 sValIleSerTyrArgGlyCysValAlaGlnLeuPheIlePheLeuAla 107
    ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
53 GACTATATCTATGTTGGGGTGCATGTA.CAACCTTTGGGGTCAATTCT 101
107 euGlyAlaThrGlyTyrLeuLeuLeuAlaValMetSerPheAspArgPhe 123
    ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 TTGGTTGACATGAGATCTTCACTTACTGTCATGCGCTATGATAGTAT 151
124 ValAlaIleCysArProLeuHisTyrSerValIleMetHisGlnArgle 140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
152 GTGGCCATTTGTAAACCTCTCCACTATATGACATATGAGCGGAGAG 201
140 uCysLeuGlnLeuAlaAlaIleSerTyrValThrGlyPheSerAsnSerV 157
    ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
202 ATGCATATAGATGTTGCTGCGAATGATGATGATGCTTCTTACATTTCTA 251
157 alrPheLeuSerThrLeuThrLeuGlnLeuProLeuCysAspProTyrVal 173
    ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
252 TTAATCCAGAGGCTTGTGTCACGCTCCATTTGTGGACCAAAATGAG 301
174 IleAspHisPheLeuGysgluValProAlaLeuLeuLysLeuSerCysVa 190
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
302 ATTGATGATCAATTTCTGTGATGATCATCTGCTGTAAGTCACTGCGCAC 351
190 lgluThr.....ThrAlaAsnGLuAlaGluLeuP 200
    ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
352 CGACACTTACATTTGCTGCTGTTGTGACAGCGCAACAGCGACCATTTG 401
200 heLeuValSerGluLeuPheHisLeuLeuProLeuThrLeuIleLeuIle 216
    ||| ||| :::: ||||| ||||| ||||| ||||| ||||| |||||
402 CATTTGGGAAGTTTTCATCTGCTGATCTCATACACAGTCATTTCCATA 451
217 SerTyrAlaPheIleValArgAlaValLeuArgIleGlnSerAlaGlu 233
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
452 TCT.....CTGAGAAGAGTCATCCGAGG 477
233 yArgGlnLysAlaPheGlyThrCysGlySerHisLeuIleValValSerL 250
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

478 CACAGCCAAAGCTCTCTCCACTTGTGATCTCATTGTCTGTTCATCA 527  
 250 eupheryserthrAlaValSerValIleuLeuInProProSerProser 266  
 528 TTTTGTGGCCCTTACTTTATGATATGCGGCTGACACTACCTTC 577  
 267 SerLysAspGlnGlyMetValSerLeuPheTyrglyIleIleAlaP 283  
 578 TCTGAGGAC.....AAGATGAGTATATTATTACACTATTATCACTCC 621  
 283 ohellLeuAsnProLeuIleTyThrLeuArgAsnLysGluValLys 298  
 622 CAGCGGAATCCCTCTAATTACCTCTAAGAAATGACAGTAAG 667

seq\_name: gb\_gss:A2909618

seq\_documentation\_block:

LOCUS A2909618 797 bp DNA GSS 05-MAR-2001  
 DEFINITION RPI-24-222G18.rj RPI-24 Mus musculus genomic clone RPI-24-222G18  
 , DNA sequence.

ACCESSION A2909618 GI:13228563

VERSION 1

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS 1 (bases 1 to 797)

TITLE Mouse BAC End Sequences from Library RPI-24

JOURNAL Unpublished (1999)

COMMENT Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhaoc@igrr.org

Clones are derived from the mouse BAC library RPI-24. For BAC

library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end

plate: http://www.tlgr.org/tlgr/bac\_ends/mouse/bac\_end\_intro.html

Plate: 222 row: G column: 18

Seq primer: SP6

Class: BAC ends.

FEATURES

Location/Qualifiers

1..797

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPI-24-222G18"

/clone\_lib="RPI-24"

/sex="Male"

/cell\_type="Spleen/Brain"

/note="Vector: pTARBA1; Site 1: BamHI; Site 2: BamHI; The library was cloned in the pTARBA1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 225 a 167 c 243 g 162 t

ORIGIN

alignment\_scores:

Quality: 497.50 Length: 266

Ratio: 2.578 Gaps: 5

Percent Similarity: 72.556 Percent Identity: 41.729

alignment\_block:

US-09-755-017-2 x A2909618/rev

Align seg 1/1 to reverse of: A2909618 from: 1 to: 797

39 IllePheGlyAsnLeuThrIleIleLeuValSerArgLeuAspThrLysle 55  
 795 CTGCTAGGCAACCTGCTCATGATGACCCATCTGAGTGAACACACACTCT 746  
 55 uHISthr.PrometTyThrPheLeuThrAsnLeuSerLeuLeuAspLeu 71  
 745 CCACACACCCCACTGATACCTCTCTGTCGCCCTCTCTCATCTCTGACATT 696  
 72 CysTyThrThrCysThrValProGlnMetLeuValAsnLeuCysSeril 88  
 695 TTCTACACCTGTGCATCATCCACGATGTTGGCTGACCTGCTCTCCAC 646  
 88 eArgLysValIleSerTyArgGlyCysValAlaGlnLeuPheIlePhe. 104  
 645 ACCTTCACCTCCATGCGCTTCTGCGCTGCGCACGACATGTTCTTCCT 596  
 105 LeuAlaLeuGlnLysAlaThrGlyLeuLeuLeuAlaValMetSerPheAs 121  
 595 TTTCACATTGCTTCACCCATCTCTTCTTACTACCTCATGGGTTATGA 546  
 121 pArgPheValAlaIleCysArgProLeuThrIleSerValIleMetHisG 138  
 545 CCGCTACGTCGTCATCTGTCCACCATGATGATCAATGTCCTCATGAGCC 496  
 138 InArgLeuCysLeuGlnLeuAlaAlaSerThrValThrGlyPheSer 154  
 495 CCGCTGTTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 446  
 155 AsnSerValThrPheLeuThrLeuThrLeuGlnLeuProLeuGlyAspPr 171  
 445 ATGGGACACATGTCGACGACGACGACGACGACGACGACGACGACGACG 396  
 171 oTyValIleAspHisPheLeuCysGluValProAlaLeuLeuLysLeus 188  
 395 CAATGAGATCCACCATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 346  
 188 eCysValGlnGlyThrThrAlaAsnGlnAlaGlu.....LeuPheLeu 201  
 345 CAGCGGAGAGAAATGACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 296  
 202 ValSerGlnLeuPheHisLeuLeuLeuLeuThrLeuIleLeuIleSer 218  
 295 ATCAGACGCGCTC.....CTGGGCTGCTTCTCTCTGATCTCTCTCAT 252  
 218 yAlaPheIleValArgAlaValLeuArgIleGlnSerAlaGluGlyArg 234  
 251 ATGCTTCATCTGCTTACCATCTTGAAGATACCATCAGCTGAGGCGCG 202  
 235 GlnLysAlaPheGlyThrCysGlySerHisLeuIleValValSerLeuP 251  
 201 CACAGGCTTCTCCACATGTCATGTCACCTCCACCTGAGTGTGTTGACA 152  
 251 eTySerThrAlaValSerValTyLeuGlnInProProSerProSerSerL 268  
 151 TTATGCTTGTCTTGTGTCATTACCTCAAGCCTTAAGGGCCCAAGTCTC 102  
 268 yAspGlnGlyMetValSerLeuPheTyrglyIleIleAlaPromet 284  
 101 TGGAGAGAGATACTCTGATGCGCATCACCACACATCTCTCACCCCTTC 52  
 285 LeuAsnProLeuIleTyThrLeuArgAsnLysGluValLys 298  
 51 CTTAGTCCCATCATCTTCACTGACGACACAGAGCTTAAG 10

seq\_name: gb\_est2:BG193339

seq\_documentation\_block:

LOCUS BG193339 762 bp mRNA EST 21-APR-2001  
 DEFINITION RST12467 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BG193339

```

VERSION      BG193339.1  GI:13715026
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    1 (bases 1 to 762)
AUTHORS      Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
              Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,
              Lerner,L., Krashinsky,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
              Veloso,N., Hess,J., Cochran,K., Lo,K., Offenbacher,J., Danzig,J.,
              and Ducar,M.
TITLE        Creation of Genome-wide Protein Expression Libraries using Random
              Activation of Gene Expression
JOURNAL      Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT      Contact: Scott J. Cain
              Athersys, Inc.
              3201 Carnegie Ave, Cleveland, OH 44115, USA
              Tel: 216 431 9900
              Fax: 216 361 9596
              Email: scaine@atersys.com
              High quality sequence stop: 551.
FEATURES     location/Qualifiers
SOURCE       1..762
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone_lib="Athersys RAGE Library"
              /cell_line="Hr1080"
              /note="See 'Creation of Genome-wide Protein Expression
              Libraries using Random Activation of Gene Expression',
              Nature Biotechnology, in press. Note that even though the
              cell type indicated is Hr1080, since a random activation
              method was used, these sequence tags are not necessarily
              expressed in Hr1080 under normal circumstances."
BASE COUNT   152 a      253 c      156 g      201 t
ORIGIN
alignment_scores:
      Quality: 493.00      Length: 243
      Ratio: 2.785      Gaps: 3
      Percent Similarity: 72.840      Percent Identity: 43.210
alignment_block:
US-09-755-017-2 x BG193339 ..
Align seg 1/1 to: BG193339 from: 1 to: 762
9  TleGlnGluPheLeuLeuGlyPheSerAspArgProTrpLeuGluPh 25
      :::::::::::::::::::::::::::::
40  GTCTCCGAGTTCCTCTGGGACCTCCAGACGCCCGCAGACAGCA 89
      ::::::::::::::::::::
25  eProLeuValValPheLeuLeuSerTyrrThValThrLeuPheGly 42
      ::::::::::::::::::::
90  TCTCTCTCTTGTCTCTCCAGCATGTCTGGCCACTCTCTGGGGA 139
      ::::::::::::::::::::
42  snLeuThrIleIleLeuValSerArgLeuAspThrIlysLeuHisThrPro 58
      ::::::::::::::::::::
140  ACCTGTCATCATCTGCGCATAGCATAGTACCTCCCGCTCCACACCC 189
      ::::::::::::::::::::
59  MetTyrrPhePheLeuThrAsnLeuSerLeuLeuAspLeuCyrrThrTh 75
      ::::::::::::::::::::
190  ATGTACTTCTCTCCAGCAACATGTCTCTTGTGACAACTGCTCTCCAC 239
      ::::::::::::::::::::
75  rCysThrValProGlnMetLeuValAsnLeuGlySerIleArgIysValI 92
      ::::::::::::::::::::
240  C...ACCGTCCCAAGATGCTGGCCAAATCAATCACTAGACAGCTAAACCA 286
      ::::::::::::::::::::
92  leSerTyrrArgGlyCysValAlaGlnPhePheLeuAlaLeuGly 108
      ::::::::::::::::::::
287  TGTCTCTCTCTGCGACGTCTCATGCAATGTATTTATTCAGAGCACTTCT 336
      ::::::::::::::::::::
109  AlatThrGluTyrrLeuLeuLeuAlaValMetSerPheAspArgPheValAl 125

```

337 GACATGGACATTTCTCTGCTGCTGATGAGCTTGTGACCCGTTTGCTG 386  
 125 ailecysarProleuNH1sYrSerValIleWethisGlnArgLeuCysL 142  
 387 CGTGGCCGCCCTTACATTTACAGCAAGAAATGACCCATGAGCTGTGG 436  
 142 euglnleuAlaIAlaSerTrpValIlnrglyPheSerAsnSerValTrp 158  
 437 CCGTGGCTGTGCTACGTGATCATGGTGTGGTCCACCTGAGTGTGCTGG 486  
 159 leuSerTrhLeuThrLeuGlnleuProleuCysAspProTyrValIleAs 175  
 487 CACACCCGTGCTGATGGCTGACTTCATCTTGTGCAACAACACCATCC 536  
 175 pHISpheLeuCysGluValProAlaLeuLeuLysLeuSerCysValGluT 192  
 537 CCACATCTTCTGCGATGTGACGCCCTCTCGAAACTCTCTGTGACACA 586  
 192 hrThAlaAsnGlnAlaGluLeuPheLeuValSerGluLeuPheHisLeu 208  
 587 CACACCTCAGTGAAGTATGATTTACTAGAGCGCCCTATGACAGATC 636  
 209 IleProleuThrLeuIleLeuIleSerTyrAlaPheIleValArgAlaVa 225  
 637 ACCCATTTCTTTGGCTCTGCTTCCTATATGACATCACCCTGGCGTGT 686  
 225 lleuArgIleGlnSerAlaGlu.GlyArgGlnLysAlaPheGlyThrCys 241  
 687 CCGTAGGGGTCAT...CCACAAGGAAAGATAAAGCCTTCTCCACTGT 733  
 242 GlySerHisLeuIleValValSerLeu 250  
 734 GGCCTCCACCTGGCTGAGGTCTCTCTC 760  
 seq\_name: gb\_gss:A2607393  
 seq\_documentation\_block:  
 LOCUS A2607393 740 bp DNA GSS 13-DEC-2000  
 DEFINITION IM0429N15R Mouse 10kb plasmid UNGC1M library Mus musculus genomic  
 clone UNGC1M0429N15 R, DNA sequence.  
 ACCESSION A2607393  
 VERSION A2607393.1 GI:11729583  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 740)  
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,  
 Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly  
 M., Rose,M., Rose,R., Stokes,R., Tingy,A., von Niederhausen,A.  
 and Wright,D., Weiss,R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
 84112, USA  
 Tel.: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert length: 10000 Std Error: 0.00  
 Plate: 0429 row: N column: 15  
 Seq primer: CACACAGAAACACCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 740.  
 Location/Qualifiers  
 1..740  
 /organism="Mus musculus"  
 /strain="C57BL/60"

/db\_xref="taxon:10090"  
 /clone="U06C1M0429N15"  
 /clone\_lib="Mouse 10kb plasmid U06C1M library"  
 /sex="Male"  
 /lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g14732114[gb]AF129072.1), a copy-number inducible derivative of plasmid RL. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 276 a 134 c 159 g 171 t  
 ORIGIN

alignment\_scores:  
 Quality: 492.00 Length: 244  
 Ratio: 2.718 Gaps: 0  
 Percent Similarity: 74.180 Percent Identity: 42.623

## alignment\_block:

US-09-755-017-2 x AZ607393/rev ..

Align seg 1/1 to reverse of: AZ607393 from: 1 to: 740

```

30 ValPheLeuIleSerThrValThrIlePheGlyAsnLeuThrIleI 46
   ::::::::::: || ::::::::::: ||::| |::| |
733 AGTTTCTAGTGTATTCTACTATCTGTCACAAATCTTGATGAT 664
   ::::::::::: || ::::::::::: ||::| |::| |
46 eLeuValSerArgLeuAspThrIleuHsrPrometTyrPhePhe 63
   ::::::::::: || ::::::::::: ||::| |::| |
683 CATTTTGATGACAAATGATGATCAGCTCACACCAATGATCTTTTC 634
   ::::::::::: || ::::::::::: ||::| |::| |
63 eutHsrAsnSerLeuLeuAspLeuGlyThrThrGlyThrValPro 79
   ::::::::::: || ::::::::::: ||::| |::| |
633 TCAGCCACCTGCTCTCTGATCTTCTGATTCACGTCAGTGTGCT 584
   ::::::::::: || ::::::::::: ||::| |::| |
80 GlnMetLeuValAsnLeuGlySerIleArgLysValIleSerTyrArg 96
   ::::::::::: || ::::::::::: ||::| |::| |
583 AGATGCTGTGATGATGATGATGATGATGATGATGATGATGATG 534
   ::::::::::: || ::::::::::: ||::| |::| |
96 yCysValAlaGlnLeuPheIlePheLeuAlaLeuGlyAlaThrGly 113
   ::::::::::: || ::::::::::: ||::| |::| |
533 CAGTGTATGACAGTCTCTGATCTTCTGATTTATAGATGCTGAGT 484
   ::::::::::: || ::::::::::: ||::| |::| |
113 euleLeuAlaValMetSerPheAspArgPheValAlaIleCysAsp 129
   ::::::::::: || ::::::::::: ||::| |::| |
483 TCCTCTGCTGACGTGATGATGATGATGATGATGATGATGATG 434
   ::::::::::: || ::::::::::: ||::| |::| |
130 LeuHsrIleSerValIleMetHsrGlnArgLeuGlyLeuAlaIle 146
   ::::::::::: || ::::::::::: ||::| |::| |
433 CTTTGTATGACAGTGTGATGATGATGATGATGATGATGATGAT 384
   ::::::::::: || ::::::::::: ||::| |::| |
383 TGGTGTATGCTGTGCGCATGATGATGATGATGATGATGATGAT 334
   ::::::::::: || ::::::::::: ||::| |::| |
163 hLeuGlnLeuProLeuGlyAspProTyrValIleAspHisPheLeu 179
   ::::::::::: || ::::::::::: ||::| |::| |
333 CATTCACCTTATGTTCTGTGGTGCATGATGATGATGATGATG 284
   ::::::::::: || ::::::::::: ||::| |::| |

```

```

180 GluValProAlaLeuLeuLysLeuSerCysValGluThrAlaAsnG 196
   ::::::::::: || ::::::::::: ||::| |::| |
283 GATATCTCTCCAGTGTAGTACTGTCTGTTCAGACACAGGTCATGT 234
   ::::::::::: || ::::::::::: ||::| |::| |
196 uAlaGluLeuPheLeuValSerGluLeuPheHisLeuIleProLeu 213
   ::::::::::: || ::::::::::: ||::| |::| |
233 TTATGTCATATTCACGCTTTTGGTTTCATTTGATGACGACCATC 184
   ::::::::::: || ::::::::::: ||::| |::| |
213 euleLeuLeuSerTyrAlaPheIleValArgAlaValLeuArgIleG 229
   ::::::::::: || ::::::::::: ||::| |::| |
183 GACTTCTATCTCTCTGATGATGATGATGATGATGATGATGATG 134
   ::::::::::: || ::::::::::: ||::| |::| |
230 SerAlaGluGlyArgGlnLysAlaPheGlyThrCysGlySerHisLeu 246
   ::::::::::: || ::::::::::: ||::| |::| |
133 TCAGTGTGCGGAGGTAAAGCTTCTCTACTGATGATGATGATGAT 84
   ::::::::::: || ::::::::::: ||::| |::| |
246 eValValSerLeuPheTyrSerThrAlaValSerValTyrLeuGln 263
   ::::::::::: || ::::::::::: ||::| |::| |
83 TGCTGTGCAATTTTTCAGGACATGCTCTTCATGATGATGATG 35
   ::::::::::: || ::::::::::: ||::| |::| |
263 roSerProSerSerLysAspGlnGlyLysMet 273
   ::::::::::: || ::::::::::: ||::| |::| |
34 GTTCTTCTACTCTCTGATGATGATGATGATGATGATGATGAT 3
   ::::::::::: || ::::::::::: ||::| |::| |

```

seq\_name: gb\_est2:BG197640

seq\_documentation\_block: 796 bp mRNA EST 21-APR-2001  
 LOCUS BG197640  
 DEFINITION RST17016 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BG197640  
 VERSION BG197640.1 GI:13719455  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 796)  
 AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,  
 Cain,S., Dahl,T., Thornton,M., Ramchandran,R., Whittington,J.,  
 Lerner,L., Krishnoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,  
 Veloso,N., Hess,J., Cochren,K., Lo,K., Offenbacher,J., Danzig,J.  
 and Ducar,M.  
 TITLE Creation of Genome-wide Protein Expression Libraries using Random  
 Activation of Gene Expression  
 JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press  
 COMMENT Contact: Scott J. Cain  
 Athersys, Inc.  
 3201 Carnegie Ave, Cleveland, OH 44115, USA  
 Tel: 216 431 9900  
 Fax: 216 361 9596  
 Email: scain@athersys.com  
 FEATURES  
 source High quality sequence stop: 544.  
 Location/Qualifiers  
 1..796  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Athersys RAGE Library"  
 /cell\_line="HT1080"  
 /note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

alignment\_scores:  
 Quality: 490.50 Length: 252  
 Ratio: 2.666 Gaps: 2  
 Percent Similarity: 73.016 Percent Identity: 42.857

alignment block:  
US-09-755-017-2 x BG197640

Align seg 1/1 to: BG197640 from: 1 to: 796

```

9  IieGlnGluPheIleuLeuGlyPheSerAspArgProThrLeuGlnph 25
   :::::::::::::::::::: ||| ::::
37 ATGTCTGAATTCATCTCTGCTGCTCTGCTGCTGCCCTCCACCCTCCACT 86
   :::::::::::::::::::: ||| ::::
25 eProLeuLeuValPheLeuIleSerTyThrValIhrlIlePheGlyA 42
   : :::::::::::::::::::: ||| ::::
87 GATGCTCTCTGCTGCTGCTGCTGATGATGCTGCTGCTGCTGCTGCTG 136
   :::::::::::::::::::: ||| ::::
42 snLeuThrIleIleLeuValSerArgLeuAspThrIleuHisThrPro 58
   |||| :::: :::: :::: :::: ::::
137 ACCGTGTCATCATGGCCACCGTCTGGAGGAGCGGCGCTCCACACGCCC 186
   :::::::::::::::::::: ||| ::::
59 MetTyRPhePheLeuThrAsnLeuSerLeuLeuAspLeuCysTyThrTh 75
   |||| :::: :::: :::: :::: ::::
187 ATGTACCTCTCTCTGCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTG 236
   :::::::::::::::::::: ||| ::::
75 rCysThrValProGlnMetLeuValAsnLeuCysSerIleArgIysVal 92
   :::: :::::::::::::::::::: ||| ::::
237 GGCATCATCCCGCGCATGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCT 286
   :::::::::::::::::::: ||| ::::
92 IeSerTyArgGlyCysValAlaGlnLeuPheIlePheLeuAlaLeuGly 108
   :::::::::::::::::::: ||| ::::
287 TCGCCTCTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 336
   :::::::::::::::::::: ||| ::::
109 AlaThrGlnTyRLeuLeuLeuValAlaMetSerPheAspArgPheVal 125
   :::: :::::::::::::::::::: ||| ::::
337 TTCACCCACTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 386
   :::::::::::::::::::: ||| ::::
125 aIleCysArgProLeuHisTySerValIleMetHisGlnArgLeuCysL 142
   :::::::::::::::::::: ||| ::::
387 CATCTGCCACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 436
   :::::::::::::::::::: ||| ::::
142 euGlnLeuAlaAlaSerTrpValThrGlyPheSerAsnSerValTrp 158
   :::::::::::::::::::: ||| ::::
437 CCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 486
   :::::::::::::::::::: ||| ::::
159 LeuSerThrLeuThrLeuGlnLeuProLeuCysAspProTyValIleAs 175
   :::::::::::::::::::: ||| ::::
487 GTGACCTCGGCCATTTCCACTGACTTTCTGTGGATTCACAGAGATCCA 536
   :::::::::::::::::::: ||| ::::
175 pHisPheLeuCysGluValProAlaLeuLeuLysLeuSerCys..... 189
   :::::::::::::::::::: ||| ::::
537 GCATTTTATGTCATGCTGCCACCTGTTGAAGTTGGCCTGGAAATA 586
   :::::::::::::::::::: ||| ::::
190 ..ValGluThrThrAlaAsnGluAlaGlnLeuPheLeuValSerGluLeu 205
   ||| :::: :::: :::: :::: ::::
587 ATGTACCACTGTGCGCTGNGCGTGGCTGTGTATGATCATGAGCACCTG 636
   :::::::::::::::::::: ||| ::::
206 PheHisLeuIleProLeuThrLeuIleLeuIleSerTyRAlaPheIleVa 222
   ||| :::: :::: :::: :::: ::::
637 .....CTGGGCTGTTTCTCTCATCTCTCTCTCTATGCTTTCATCGT 680
   :::::::::::::::::::: ||| ::::
222 IArgAlaValLeuArgIleGlnSerAlaGlnGlyArgGlnLysAlaPheG 239
   :::::::::::::::::::: ||| ::::
681 GCGCGACATCTGAAGATCCCTTCTGCTGAAGTCGGAACAAGCCCTTCT 730
   :::::::::::::::::::: ||| ::::
239 IyThrCysGlySerHisLeuIleValValSerLeuPheTySerThrAla 255
   :::::::::::::::::::: ||| ::::
731 CCACCTGTGCTCTACCTTAATGGTATGTCATATGCTATGCTTTCGT 780
   :::::::::::::::::::: ||| ::::
256 ValSer 257
   ::::|||
781 CTGTCA 786
```







```

172 AGTGCATGCTTAACACACCCCATGTAAGTCTTTCTGGCAACTGTGTCCT 221
      ::::::::::::::::::::
68 uLeuAspSerCysGlyThrThrCysThrValProGlnMetLeuValAsnL 85
      ::::::::::::::::::::
222 CCTGAGATGTGGCTATACTCTGCTGCTGATACCCAGATGAGTGTGACAGCC 271
      ::::::::::::::::::::
85 eucysSerlleargLysValIleSerTyArgGlyCysValAlaGlnLeu 100
      ::::::::::::::::::::
272 TTGTGAGTGAAGCCCAAGAGATCTCTGGGAGGAGTGTCCACACGATG 321
      ::::::::::::::::::::
102 PheIlePheLeuAlaLeuGlyAlaThrGlyTyLeuLeuLeuAlaValMe 118
      ::::::::::::::::::::
322 TTTTCTTTTCCGATTTTGGTATTAAGTATGAGTGTGCTGCTTTGGCAGCCAT 371
      ::::::::::::::::::::
118 tSerPheAspArgPheValAlaIleCysArgProLeuHisTySerValI 139
      ::::::::::::::::::::
372 GGCCTTTGACCGCTGCATGCGCATATGCTGCCACATCCATAGCAACC 421
      ::::::::::::::::::::
135 LeuMetHisGlnArgLeuCysLeuGlnLeuAlaAlaIleSerThrValThr 151
      ::::::::::::::::::::
422 GAATGACGTGCGAGAGTATGTGCCATTTGGCAATGTTTCATGGGAAATG 471
      ::::::::::::::::::::
152 GlyPheSerAsnSerValThrPheSerThrIleuThrLeuGlnLeuProle 168
      ::::::::::::::::::::
472 GCAATGCATAGTAAGTCTGGCAAAACCAATTTATTTTCTGCTTGAATTT 521
      ::::::::::::::::::::
168 uCysAspProTyrAlaIleAspHisPheLeuCysGluValProAlaLeu 185
      ::::::::::::::::::::
522 CTGTGAGAACCCGTGAATAAGAACACTTCTGCTGAGCTTCCACACTCTCC 571
      ::::::::::::::::::::
185 euLysLeuSerCysValGluThrThrAlaAsnGlnAlaGluLeuPheLeu 201
      ::::::::::::::::::::
572 TGGCATTGCGCTGTGGAGATACATCCAAAGAGAGGCTGCATCTTTGTG 621
      ::::::::::::::::::::
202 ValSerGluLeuPheHisLeuIleProLeuThrIleuIleLeuIleSerTy 218
      ::::::::::::::::::::
622 GTAGACATGCTCTGCATATCTAGACCAATTTTGGCATCATATTATCTTA 671
      ::::::::::::::::::::
218 rAlaPheIleValrAlaValAlaLeuArgIleGlnSerAlaGluArg 235
      ::::::::::::::::::::
672 TGTCAAAATTTCTCATGTGCAGTGTCTGCTGATGCTTCCACTGAGGGCGCC 721
      ::::::::::::::::::::
235 lInLysAlaPheGlyThrCysGlySerHisLeuIleValValSerLeuPhe 251
      ::::::::::::::::::::
722 ATTAAGCTCTTCCACCGCTGTGCTGCTACGTAAGCTGTGATACACTTTT 771
      ::::::::::::::::::::
252 TyrSerThrAlaValSerValTyTyLeuGlnProProSerProSerSerTy 268
      ::::::::::::::::::::
772 TATGGCTCAGCATTTATTAACCTATTGTGAGGCCCAAGTCTAGCCACTACCC 821
      ::::::::::::::::::::
268 saSpGlnGlyLysMetValSerLeuPheTyGlyIleIleAlaPrometL 285
      ::::::::::::::::::::
822 AGGAATGACCAAAATTTGTGGCCCTCTTTACACAGATGAGATCATCATG 871
      ::::::::::::::::::::
285 euAsnProLeuIleTyThrThrLeuArgAsnLysGlnValLysGluGlyPhe 301
      ::::::::::::::::::::
872 TGAACCTTATCATGTATTAAGTTAAGGAACAGGAAGTCAAGCAGCAGCTG 921
      ::::::::::::::::::::
302 LysArg 303
      ::::::::::::::::::::
922 AGAAGA 927

```

ADDRESSEE: Leydig, Voigt & Mayer, Ltd.  
STREET: Two Prudential Plaza, Suite 4500  
City: Chicago  
STATE: IL  
COUNTRY: US  
ZIP: 60601-6780  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/748,506  
FILING DATE: 08-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/033,751  
FILING DATE: 09-NOV-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
REFERENCE/DOCKET NUMBER: 74940  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-616-5600  
TELEFAX: 312-616-5700  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 966 base pairs  
type: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-748-506-7

```
alignment_scores:
  Quality: 652.50      Length: 314
  Ratio: 2.825         Gaps: 2
Percent Similarity: 73.567      Percent Identity: 42.999
```

```
alignment_block:
US-09-755-017-2 x US-08-748-506-7 .
```

Align seg 1/1 to: US-08-748-506-7 from: 1 to: 966

[illegible]

```

135 lemeHISGLArgLeucysLeuGlnLeuAlaAlaSerThrValThr 151
422 GAATAGTGTGGGGATGTGCGCTATTGGCAATTTGTCAGGGGATG 471
152 GlyPheSerAsnSerValThrPheSerThrLeuThrGlnLeuProle 168
472 GGATTCATAGTAGTGTGGGAGACACCAATTTATTTCCTTGAACCT 521
168 uCysAspProTyrValIleAspHisPheLeuCysGlnValProAlaLeu 185
522 CTGTGACCCCTGTGAGATAGACCACTTCTTGTGACCTTCACCTCTCC 571
185 euLysLeuSerCysValGluThrThrAlaAsnGlnAlaGluLeuPheLeu 201
572 TGGCACTTGGCCGTGTGATCATCCCAAAAGAGGCTGCCATCTTTGTG 621
202 ValSerGluLeuPheHisLeuIleProLeuThrLeuIleLeuIleSerTy 218
622 GCAGCAGTGTCTGCATATTAGTCCATTTTACTGATCATTTCTTCTTA 671
218 talapheIleValArgAlaValLeuArgIleGlnSerAlaGluGlyArg 235
672 TGTCAGAATTTCTGTTGGAGTGTGATGCTGCCTTCACCTGAGGGGCGCC 721
235 InLysAlaPheGlyThrCysGlySerHisLeuIleValValSerLeuPhe 251
722 ATAAAGCTCTCTCTACCTGTTCATCTCACCTGTGTAGTCAACCTCTTC 771
252 TyrSerThrAlaValSerValThrLeuGlnProProSerProSerSerTy 268
772 TATGGCTCAACATCTGCCACCTATTGAGGTCCAAAGTCAAGCCACCTACC 821
268 sasplnGlyLysMetValSerLeuPheThrGlyIleIleAlaProMetL 285
822 AGGAGTGGCAACACCTTGGCCCTCTCTATACATCAGTACATCATCATC 871
285 euAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGluGlyPhe 301
872 TGAATCCCATCTCATACACCTTAAGGAAACAGTAAGGAGTGAAGGCTG 921
302 LysArgLeuValAla..... ArgValPheLeuIleLysLys 313
922 ACAAGAACTCTGGGCTGAGAAAGTTCTGACAAATGAAGAAG 963
seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-467-948A-1
seq_documentation_block:
: Sequence 1, Application US/08467948A
: Patent No. 5998164
: GENERAL INFORMATION:
: APPLICANT: LI, YI
: APPLICANT: CAO, LIANG
: APPLICANT: NI, JIAN
: APPLICANT: GENTZ, REINER
: APPLICANT: BULT, CAROL J.
: APPLICANT: SUTTON III, GRANGER G.
: APPLICANT: ROSEN, CRAIG A.
: TITLE OF INVENTION: Polynucleotides Encoding Human G-Protein
: TITLE OF INVENTION: Coupled Receptor GPR2
: NUMBER OF SEQUENCES: 30
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
: STREET: 1100 NEW YORK AVE., NW, SUITE 600
: CITY: WASHINGTON
: STATE: DC
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: FLOPPY DISK
: COMPUTER: IBM PC COMPATIBLE

```

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PATENTIN RELEASE #1.0, VERSION #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,948A
CLASSIFICATION: 435
FILING DATE: 06-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04079
FILING DATE: 30-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.1140003/EKS/KLM
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1713 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
MOLECULE TYPE: both
FEATURE:
NAME/KEY: CDS
LOCATION: 116..1003
US-08-467-948A-1

```

## alignment\_scores:

```

Quality: 652.50 Length: 295
Ratio: 2.900 Gaps: 3
Percent Similarity: 76.271 Percent Identity: 45.424

```

## alignment\_block:

```

US-09-755-017-2 x US-08-467-948A-1 ..

```

```

Align seg 1/1 to: US-08-467-948A-1 from: 1 to: 1713

```

```

5 AsnAspSerIleIleGlnGluPheIleLeuLeuGlyPheSerAspArgP 21
1137 AATCAGACAAATGGTCAACAGATTCCTCTACCTGATTCCTCTGAGCC 186
21 crrPheGluPheProLeuLeuValAlaPheLeuIleSerTyrThrVal 38
187 AAGGATTCAGATGCTCTCTTGGGCTCTCCCTGTTCTATGTTCTCA 236
38 hrIlePheGlyAsnLeuThrIleIleLeuValSerArgLeuAspThrLys 54
237 CCTGCTGGGGAATGGGACCATCTTGGGCTCATCTCAGTGCATCCAGA 286
55 LeuHisThrProMetTyrPhePheLeuThrAsnLeuSerLeuLeuAspLe 71
287 CTCACACACCCCATCTACTTCTCTCTCTCACACCTGGCGGTGCAACAT 336
71 uCysTyrThrThrCysThrValProGlnMetLeuValAsnLeuCysSerI 88
337 CGCTATGGCTGCACACACAGTGCCTGATGCGGAGAACCTCTCCATC 386
88 leArgLysValIleSerTyrArgGlyCysValAlaGlnLeuPheIlePhe 104
387 CAGCCAAAGCCATCTCTTGGCTGCTGATGACATGACACTTCTCTTT 436
105 LeuAlaLeuGlyAlaThrGluTyrLeuLeuLeuAlaValMetSerPheAs 121
437 TTGAGTTTGCATATCATGATGATGCTGTTGGTGTGATGCTGATGCTCA 486
121 parGpHeValAlaIleCysArgProLeuHisTyrSerValIleMetHisG 138
487 TCGGTAGTGGCCATCTGCCACCTCTCCGATATTTTCATCATCATGACCT 536
138 InArgLeuCysLeuGlnLeuAlaAlaAlaSerThrValThrGlyPheSer 154

```

```

537 GGAAGTCTGCATCACTCTGGGACATCTTCTGGACATGTGCTCCCTC 586
155 AsnSerValTrpLeuSerThrLeuThrLeuGlnLeuProLeuGlyAspPr 171
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
587 CTGGCTATGTCATGTGACCTCATCTCACTAGACTGCCCTTTGTGGGCC 636
171 oTyValIleAspHisPheLeuGlyGluValProAlaLeuLeuLysLeus 188
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
637 TCCTGAATCAACCACTTCTTGTGTAATCTGTCTGCTGCTCAGGCTGG 686
188 eTyValGluThrThrAlaAsnGlnAlaLeuLeuPheLeuValSerGlu 204
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
687 CCGTGTGCTGATACCTGGCTGACACAGGTGCTCATCTTTGAAGCTGCATG 736
205 LeuPheHisLeuLeuProLeuThrLeuLeuLeuLeuLeuSerTyraIle 221
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
737 TTGATCTGTGGGACCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 786
221 eValArgAlaValLeuArgIleGlnSerAlaGluGlyArgGlnLysAlaP 238
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
787 CTGGGGGGGACATCTGAGATCCAGTCTGGGAGGGCCGAGAAAGGCT 836
238 heGlyThrGlySerHisLeuLeuValValSerLeuPheTyrrSerThr 254
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
837 TCTCCACCTGCTCTCCACCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 886
255 AlaValSerValTyrrLeuGlnProProSerProSerSerLysAspGln 271
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
887 GCATCGTCACTGATGACATGGCCCTAAGTCCCGCCATCTAGAGACGCA 936
271 yLysMetVal.SerLeuPheTyrrGlyIleIleAlaPro..MetLysn 286
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
937 GAAGGTCCTTTTCTTATTTTACAGTCTTCAACCCCGATGCTTAAAC 986
287 ProLeuIleTyrr.ThrLeuArgAsnLysGlu 296
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
987 CCCGTATTTTACACCTGAGGATGATGAG 1017

```

## seq\_name: /cgn2\_6/plodata/2/lna/6A\_COMB.seq:us-08-467-947A-1

```

seq_documentation_block:
: Sequence 1, Application US/08467947A
: Patent No. 6090575
: GENERAL INFORMATION:
: APPLICANT: LI, YI
: APPLICANT: CAO, LIANG
: APPLICANT: NI, JIAN
: APPLICANT: GENTZ, REINER
: APPLICANT: BULT, CAROL J.
: APPLICANT: SUTTON III, GRANGER G.
: APPLICANT: ROSEN, CRAIG A.
: TITLE OF INVENTION: Polynucleotides Encoding Human G-Protein
: TITLE OF INVENTION: Coupled Receptor GPR1
: NUMBER OF SEQUENCES: 30
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
: STREET: 1100 NEW YORK AVE., NW, SUITE 600
: CITY: WASHINGTON
: STATE: DC
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: FLOPPY DISK
: COMPUTER: IBM PC COMPATIBLE
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PATENTIN RELEASE #1.0, VERSION #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/467,947A
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/04079
: FILING DATE: 30-MAR-1995

```

```

ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488,1140002/EKS/KLM
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1713 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 116..1003
US-08-467-947A-1

```

```

alignment_scores:
Quality: 652.50 Length: 295
Ratio: 2.900 Gaps: 3
Percent Similarity: 76.271 Percent Identity: 45.424

```

## alignment\_block:

US-09-755-017-2 x US-08-467-947A-1 ..

Align seg 1/1 to: US-08-467-947A-1 from: 1 to: 1713

```

5 AsnAspSerIleIleGlnGluPheIleLeuLeuGlyPheSerAspArgPr 21
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
137 AATCAGACATGTGTCACAGAGTTCCTCCCTAGTGGATTTCTCCGGGCC 186
21 oTyrrLeuGluPheProLeuLeuValValPheLeuIleSerTyrrThrVal 38
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
187 AAGGATCAGATGCTCTCTTGGGCTCTTCTCCCTGTTCTATGCTTCA 236
38 hTrpPheGlyAsnLeuThrIleIleLeuValSerArgLeuAspPheIly 54
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
237 CCTGCTGGGGAATGGACCATCTCGGGCTCATCTCAGCTGACCTCAGA 286
55 LeuHisThrProMetTyrrPhePheLeuThrAsnLeuSerLeuLeuAsp 71
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
287 CTCACACCCCATGTAATCTTCTCTCTCACACTGGCCCTGTCACAAAT 336
71 uCysTyrrThrThrCysThrValProGlnMetLeuValAsnLeuCysSer 88
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
337 CGCTATGCTGTCACACACTGCCGCCAGATGCTGTGAACCTCTGCATC 386
88 LeuArgValIleSerTyrrArgGlyCysValAlaGlnLeuPheIlePhe 104
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
387 CACCCAGCCCATCTCTTGGCTGTGCAATGACACTAGACTTCTCTTT 436
105 LeuAlaLeuGlyAlaThrGluTyrrLeuLeuLeuAlaValMetSerPhe 121
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
437 TTGAGTTTGCATCATCTGATGATGCTCTGCTGCTGCTGCTGCTGCTG 486
121 PArgPheValAlaIleCysArgProLeuHisTyrrSerValIleMetHis 138
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
487 TCGGTAGTGGCCATCTGACCCCTCCGATTTTCATCATCATGACCT 536
138 LnaArgLeuGlyLeuGlnLeuAlaAlaValSerTrpValTrgIlyPheSer 154
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
537 GGAAGTCTGCATCACTCTGGGACATCTTCCGACATGCTGCTCCCTC 586
155 AsnSerValTrpLeuSerThrLeuThrLeuLeuLeuLeuLeuSerTyra 171
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
587 CTGGCTATGTCATGTGACCTCATCTCACTAGACTGCCCTTTGTGGGCC 636
171 oTyrrValIleAspHisPheLeuGlyGluValProAlaLeuLeuLysLeus 188
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
637 TCCTGAATCAACCACTTCTTGTGTAATCTGTCTGCTGCTCAGGCTGG 686

```



```

1  NUMBER OF SEQUENCES: 2
2  CORRESPONDENCE ADDRESS:
3  ADDRESSEE: Smlthklne Beecham Corporation
4  STREET: 709 Swedeland Road
5  CITY: King of Prussia
6  STATE: PA
7  COUNTRY:
8  ZIP: 19406
9  COMPUTER READABLE FORM:
10 MEDIUM TYPE: Diskette
11 OPERATING SYSTEM: DOS
12 SOFTWARE: FASTSEQ for Windows Version 2.0
13 CURRENT APPLICATION DATA:
14 APPLICATION NUMBER: US/08/827,291A
15 FILING DATE: 28-MAR-1997
16 CLASSIFICATION: 435
17 PRIOR APPLICATION DATA:
18 APPLICATION NUMBER:
19 FILING DATE:
20 ATTORNEY/AGENT INFORMATION:
21 NAME: King, William T
22 REGISTRATION NUMBER: 30,954
23 REFERENCE/DOCKET NUMBER: GP50001
24 TELECOMMUNICATION INFORMATION:
25 TELEPHONE: 610-270-5015
26 TELEFAX: 610-270-5090
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995

```

Quality:	648.00	Length:	301
Ratio:	2.830	Gaps:	2
Percent Similarity:	76.080	Percent Identity:	43.189

```

align seg 1/1 to: US-08-827-291A-1 from: 1 to: 1290
alignment_block:
US-09-755-017-2 x US-08-827-291A-1 ..

```

[illegible][illegible]

```

seq_name: /cogn2_6/ptodata/2/ina/6a_COMB.seq:us-08-748-
seq_documentation_block:
? Sequence 8, Application US/08748506
? Patent No. 6159707
? GENERAL INFORMATION:
? APPLICANT: Ronnett et al.
? TITLE OF INVENTION: NOVEL SPERM RECEPTORS
? NUMBER OF SEQUENCES: 31
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Leydig, Volt & Mayer, Ltd.
? STREET: Two Prudential Plaza, Suite 4900
? CITY: Chicago
? STATE: IL
? COUNTRY: US
? ZIP: 60601-6780
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:

```

```

152 CTGTGACCCCTGTAGATAGACACCTTCTTGATCTTCACCTCC
185 euLysLeuSerCysValGluThrThrAlaAsnGluAlaGluLeuPheLeu
572 TGGCAGCTTGGCTGGTGATACATGCCAAATTGAGGCGCCACTCTTGTG
202 ValSerGluLeuPheHisLeuIleProLeuThrLeuIleLeuIleSery
622 GAGTGTGTCCTCGCATCTACGCCCTTTTGGCTGATTTATCTTA
218 rAlaPheIleValArgAlaValIleuArgIleGlnSerAlaGluIlyArg
672 TGTCAAGATTCTGTTGCAAGCTGCTGTGATGCTTCACTGAGGGGGCC
235 IuLysAlaPheGlyThrCysGlySerHisLeuIleValSerLeuPhe
722 ACAAAAGCCCTTTCACACGCTTCTCCACACCTGTAAGTACACTCTTT
252 TyrSerThrAlaValSerValIlyLeuGlnProProSerProSerIly
772 TATGGCTCAGAGTCTGTACCTATTGTAGGCCCTAAGCTTACCCACTCAC
268 sasplngIySmetValSerLeuPheTyrGlyIleIleAlaProMetL
822 AGCAATGGACAACACTTGGCCCTTCTTACACAGCAGCATGCATCCATGT
285 euAspProLeuIleTyrThrLeuArgAsnIySgluValIySgluIyPhe
872 TGAACCCATCATCTATATAGTTTAAGAGCAACAGATGTCAAGCAGCAGCTG
302 IysArgLeuValAla.....ArgValPheLeuIleIySlys 313
922 AGAAGATTCTGGCCCTGAAAAAATCTGTGCATTAATAATAG 963

seq_name: /cgn2_6/prodata2/lna/orA_COMB.seq:US-08-748-506-6

seq_documentation_block:
- Sequence 6, Application US/08748506
- Patent No. 6159707
- GENERAL INFORMATION:
  APPLICANT: Ronnett et al.
  TITLE OF INVENTION: NOVEL SPERM RECEPTORS
  NUMBER OF SEQUENCES: 31
  CORRESPONDENCE ADDRESS:
  ADDRESSEE: Leydig, Voigt & Mayer, Ltd.
  STREET: Two Prudential Plaza, Suite 4900
  CITY: Chicago
  STATE: IL
  COUNTRY: US
  ZIP: 60601-6780
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/748,506
  FILING DATE: 08-NOV-1996
  CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 60/033,751
  FILING DATE: 09-NOV-1995
  CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
  REFERENCE/DOCKET NUMBER: 74940
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: 312-616-5600
  TELEFAX: 312-616-5700
  INFORMATION FOR SEQ ID NO: 6:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 966 base pairs
  type: nucleic acid

```

STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-748-506-6

alignment\_scores:  
 Quality: 636.50 Length: 302  
 Ratio: 2.893 Gaps: 1  
 Percent Similarity: 72.848 Percent Identity: 43.709

alignment\_block:

US-09-755-017-2 x US-08-748-506-6 ..

Align seg 1/1 to: US-08-748-506-6 from: 1 to: 966

```

3 TrpValAsnAspSerile...IleGlnGluPheIleuLeuGlyPhe 18
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
22 TGCAGAGAGATAGTGTCTGTCACAAAGCTTTCATTTGCCAGCTTCTC 71
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
18 rAspArgProTrpLeuGluPheProLeuValValPheLeuIleSerT 35
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
72 TGAGGTCCCTGGAGAAATGCTTCCCTGTCACCCCTCATCTCTCATGT 121
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
35 yTrThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArgLeu 51
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
122 TCTTAGATCACTACAGAGAAATCTCTCATAGCCCTTGCTGTTGTACC 171
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
52 AspThrIysLeuHisThrProMetLysrPhePheLeuThrAsnLeuSer 68
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
172 AGTCATCTCTACACACACCCCATGCTTCTTCTGCGCAACTGTCTCT 221
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
68 uLeuAspLeuGlyThrThrIleGlyThrValProGlnMetLeuValAsn 85
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
222 CCTGGAGATTGGCTATAGTCTGCTGTCATACCCAGATGCTGCGAGGC 271
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
85 euCysSerIleArgLysValIleSerTyraArgLysValAlaGlnLeu 101
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
272 TTGTGATGTAGGCCAGAGAGATCTCTAGGAGGATGTGCCACACATG 321
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
102 PheIlePheLeuAlaLeuGlyAlaThrGluTrpLeuLeuAlaValMe 118
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
322 TTTTCTTTCACATTTTGGCATTAAGTACGAGTGGCTATTTGGCAGCAT 371
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
118 tSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSerVal 135
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
372 GGCCTTGGACCGCTGATGGCATATGCTCCCATCTCCACTATGCAACC 421
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
135 leMetHisGlnArgLeuGlyLeuGlnLeuAlaAlaIleSerTrpValThr 151
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
422 GAATGAGTCTGTGAGGTATGTGCCATTTGGCAATTTGTTTCATGGGAA 471
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
152 GlyPheSerAsnSerValTrpLeuSerThrLeuThrLeuGlnLeuPro 168
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
472 GGAATGCACTAGTGTGCTGGAGACAGCAATTTNATTTNCTTGAAGCT 521
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
168 uCysAspProTyraIleAspHisPheLeuGlyValProAlaLeu 185
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
522 CTGTGGACCTTGTGAGATAGCACTTCTGTGACCTTCCACCTTCTCC 571
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
185 euLysLeuSerCysValGluThrThrAlaAsnGluAlaGluLeuPhe 201
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
572 TGGACATGCTGCTGTGATATACATCCCAAAACAGAGCTGCCATCTTGT 621
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
622 GCAGCAATCCTCTATATATCTAGTCCATTTTGGTATCTTATATCTTA 671
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
202 ValSerGluLeuPheHisLeuIleProLeuThrLeuLeuIleSerT 218
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
622 GCAGCAATCCTCTATATATCTAGTCCATTTTGGTATCTTATATCTTA 671
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
218 rAlaPheIleValAlaGlyAlaValLeuArgIleGlnSerAlaGluGly 235
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
672 TGTGAGATTTCTGTTGATGATGATGATGATGATGATGATGATGATG 721
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
235 InLysAlaPheGlyThrCysGlySerHisLeuIleValSerLeuPhe 251
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||

```

```

722 ATTAAGCTCTTCCACCTGTTCCACACCTACTACTAGTACACTCTTT 771
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
252 TyrSerThrAlaValSerValTyrLeuGlnProProSerProSerSer 268
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
772 TATGCTCTGTGCTTACCTATTTGAGGCCCAAGCTAGCCACCTACC 821
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
268 sAspGlnGlyLysMetValSerLeuPheTyrglyIleIleAlaProMet 285
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
822 AGGAATGACAAACTTGGCCCTTCTACACAGCAGTACATCCATGTC 871
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
285 euAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGluGlyPhe 301
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
872 TGAACCTATCATCTACAGTCTAGAGAACAGAGATCAAGCAGCACTG 921
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
302 LysArg 303
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
922 AGAAGA 927
  ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||

```

seq\_name: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-748-506-9

seq\_documentation\_block:

Sequence 9, Application US/08748506

Patent No. 6159707

GENERAL INFORMATION:

APPLICANT: Konnett et al.

TITLE OF INVENTION: NOVEL SPERM RECEPTORS

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESSEE: Leydig, Volt & Mayer, Ltd.

STREET: Two Prudential Plaza, Suite 4900

CITY: Chicago

STATE: IL

COUNTRY: US

ZIP: 60601-6780

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/748,506

FILING DATE: 08-NOV-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/033,751

FILING DATE: 09-NOV-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

REFERENCE/DOCKET NUMBER: 74940

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-616-5600

TELEFAX: 312-616-5700

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 984 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-748-506-9

alignment\_scores:  
 Quality: 587.00 Length: 319  
 Ratio: 2.552 Gaps: 3  
 Percent Similarity: 72.100 Percent Identity: 36.364

alignment\_block:  
 US-09-755-017-2 x US-08-748-506-9 ..

Align seg 1/1 to: US-08-748-506-9 from: 1 to: 984



```

1 MetAsnTrpVal.....AsnAspSerIleIleGlnI 11
1 ATGGAATGCTAATGACAGAGACAGACAGAAATGCTTGGCTCGA 50
11 uPheIleuLeuGlyPheSerAspArgProTrpLeuGluPheProLeu 28
51 GTTCATCTTGAGGGTACCTGTGGCCGAGACACTGAAGATCCTCTCT 100
28 euValValPheLeuIleSerTrpThrValThrIlePheGlyAsnLeuThr 44
101 TCCTACTGACCTGTGGCTACTTGGCCCTCCATGGCAACATGCTC 150
45 IleIleLeuValSerArgLeuAspThrLysLeuHisThrPrometLyrp 61
151 ATATTACCATCATCTGTGTGACACCGACTGACAGCGCCATGACTT 200
61 ePheLeuThrAsnLeuSerLeuAspLysCysTrpThrThcCysThrV 78
201 CTTTCTGACACTCTCTCTTTGTGAGTGTGTTTATACACTGCTA 250
78 aLProGlnMetLeuValAsnLeuCysSerIleArgLysValIleSerTrp 94
251 TCCCCAGAGCTCCACCATCATCTGTGACAGGAGCAAGATTCCTT 300
95 ArgGlyCysValAlaGlnLeuPheIlePheLeuAlaLeuGlyAlaThrG 111
301 GGGGTCTGCTTCTACAGGCTTGTCTATCTGTGCGGGGCAACAG 350
111 uTyrlleuLeuAlaValMetSerPheAspArgPheValAlaIleCysA 128
351 TTTTTCCTTTGGCTGCTTATCCCTGACCGCTTCTGGCCATCTGCA 400
128 rPProLeuIstLysSerValIleMetHisGlnArgLeuCysLeuGlnLeu 144
401 AACCTTACATTTCCACCATCATGACCCCAAGATGCTCTCTCTTC 450
145 AlaAlaIleSerTrpValThrGlyPheSerAsnSerValTrpLeuSerTh 161
451 GTTACTGCTGTTATTATTGGGCTTC.....CTCTCATGGCCAG 491
161 rLeuThrLeuGlnLeuPro.....LeuCysAspProTrpValIleA 175
492 TCCAGTGTGATGCTTCCAGACATTTACTGTGCTCCAAACATTAATTC 541
175 sPHisPheLeuCysGluValProAlaLeuLeuLysLeuSerCysValGlu 191
542 CTACACTTTTCTGTGATTTTGGACCACTGGCAATCTCTCTCTTCAGAA 591
192 ThrThrAlaAsnGluAlaGlnLeuPheLeuValSerGluLeuPheHisLe 208
592 ACCAGGCTATTTGAGATGCTTTTAACTTGGCTGATTTGCTTTT 641
208 uIlePheLeuThrLeuIleLeuIleSerTrpAlaPheIleValAlaVal 225
642 TGCCTTCCTTCTTATGACCATCTTTCATGACCAATATAGTAGTCACCA 691
225 aLeuArgIleGlnSerAlaGlnGlyArgGlnLysAlaPheGlyThrCys 241
692 TACTGAGACTCCCTTCAGCCAGGAGGAGAGAGCTTTTTCACCTGTC 741
242 GlySerHisLeuIleValValSerLeuPheTrpSerThrAlaValSerVa 258
742 TCCTTCATCTCATTTGCTCTCTTAATGATGACGCTGCTCATTTAT 791
258 lTyrlleuGlnProSerProSerSerLysAspGlnGlyLysMetValS 275
792 ATACCGAAGCCAAAGCAGAGAGAGAGAGACCAACAGAGAGAGCTG 841
275 eLLeuPheTrpGlyIleIleAlaProMetLeuAsnProLeuIleTrp 291
842 CTCTTGTGAACATGTTGTGACACCCCTTGTGAACCTGTGATCATCTAC 891
292 LeuArgAsnLysGluValLysGlnGlyPheLysArgLeuValAlaArgVa 308

```

```

892 CTGCGCAACAGACAGCTCCACAGGCTCTCAGGAGTCTCTGTCACAGCT 941
308 lPheLeu 310
942 TCAATTA 948

seq_name: /cgn2_6/plodata/2/ina/6A_COMB.seq:us-08-988-876-2
seq_documentation_block:
; Sequence 2, Application US/08988876
; Patent No. 6063596
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yee, Henry
; TITLE OF INVENTION: G PROTEIN COUPLED RECEPTORS ASSOCIATED
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,876
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0441 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1828 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSNOT01
; CLONE: 364702
; US-08-988-876-2

alignment_scores:
Quality: 417.50 Length: 296
Ratio: 2.119 Gaps: 5
Percent Similarity: 66.554 Percent Identity: 37.500

alignment_block:
US-09-755-017-2 x US-08-988-876-2 ..
Align seg 1/1 to: US-08-988-876-2 from: 1 to: 1828

12 PheIleuLeuGlyPheSerAspArgProTrpLeuGluPheProLeu 28
|||||
814 TTCTCTCTCTGAGAGTCTCAGGAGGATCCAGAACTGACGACGCTTGG 863

```

```

28 uValValPheLeuIleSerThrValThrIlePheGlyAsnLeuThrI 45
   ::::::::::::::::::::
864 TGGGCTGTCTCTGACATGCTGTCACGGTGTGGGAACTGTGCTA 913
45 IleIleuValSerArgLeuAspThrLysLeuHisThrPrometThrPhe 61
   ::::::::::::::::::::
914 TCATCCGCGGCATCAGCCCTGACTCCACCTCCACACCCCATGTACTTC 963
62 PheLeuThrAsnLeuSerIleuLeuAspLeuCysThrThrLysThrVal 78
   ::::::::::::::::::::
964 TTCTCTCCCAACCTGTCTTGGCTCCATCGGTTTCACCTCCACACGAGT 1013
78 1ProGlnMetLeuValAsnLeuCysSerIleArgLysValIleSerThr 95
   ::::::::::::::::::::
1014 CCCCAAGATGATGTGTGACATCCAGTCTCACAGACAGATCATCTCCATG 1063
95 rglGlyCysValAlaGlnLeuPheIlePheLeuAlaLeuGlyAla...Thr 110
   ::::::::::::::::::::
1064 CAGGCTGCCCTGACATCAGATGTCTCTTCCATTTTGGAGGCATGGAA 1113
111 GluThr LeuLeuLeuAlaValMetSerPheAspArgPheValAlaIleC 127
   ::::::::::::::::::::
1114 GAGAGACATCTCTCTGATGTGATGGCTTATGACTGGTTGTAGCCATCT 1163
127 ySarGProLeuHisThrSerValIleMetHisGlnArgLeuCysLeuGln 143
   ::::::::::::::::::::
1164 GTCAACCGGATATGATTCACAC.ATCATGAACCGGTGTCTGTGGCTTT 1212
144 LeuAlaAlaAlaSer...TrpValThrGlyPheSerAsnSerValTrpLe 159
   ::::::::::::::::::::
1213 CTAGTTTGTGTGCTTTTCTTTTCTTCATGCTTTTACACTCCACAGCTGA 1262
159 uSerThrLeuThrLeuGlnLeuProLeuCysAspProThrValIleAspH 176
   ::::::::::::::::::::
1263 CAACCTATATGCTTCTACAGTACGCTCTTCAGAGATGTGGAATTTCTTA 1312
176 1sPheLeuCysGluValProAlaLeuLeuLysLeuSerCysValGluThr 192
   ::::::::::::::::::::
1313 ATTTCTTCTGTGACCTTCTTCACCTCTCCCATCTTGCATGTGTGGACACC 1362
193 ThrAlaAsnGluAlaGlnLeuPheLeuValSerGluLeuPheHisLeuI 209
   ::::::::::::::::::::
1363 TTCACCATTAACATATCATGATTAATTCCTGCTGCCATTAATTTGGTTTCT 1412
209 eProLeuThrLeuIleLeuIleSerThrAlaPheIleValArgAlaValI 226
   ::::::::::::::::::::
1413 TCCCATCTCAGGACCTTTCTCTTACTGTAATAATTTCTTCCCTCCATTC 1462
226 euArgIleGlnSerAlaGluGlyArgGlnLysAlaPheGlyThrCysGly 242
   ::::::::::::::::::::
1463 TGAGGGTTTCATCATCAGTGGGAAATTAACCT.TTCTCCACCTGTGGG 1511
243 SerHisLeuIleValVal.SerLeuPheThrSerThrAlaValSerValT 259
   ::::::::::::::::::::
1512 TTCACCTCTGACGTTGTTGCT.....GGAGGCT 1540
259 yIleGlnProProSerProSerSerLysAspGlnGlyLysMetValSer 275
   ::::::::::::::::::::
1541 ACCTCGGTAGATGTGTATCTTCCCGGAAAGAGATGACAGTGGCTCTCA 1590
276 LeuPheThrGlyIleIleAlaPrometLeuAsnProLeuIleThrIle 292
   ::::::::::::::::::::
1591 GTGATGTACACGGTGTGTCACCCCATGCTGACCCCTTCATGTATACACCT 1640
292 uArgAsnLysGluValLysGluGlyPheLysArg 303
   ::::::::::::::::::::
1641 GAGAAACAGGATATGAAGATGCTCCGCGCGG 1674
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-465-980-1
seq_documentation_block:
; Sequence 1, Application US/08465980
; Patent No. 5756309

```

```

GENERAL INFORMATION:
APPLICANT: Soppet, Daniel R.
APPLICANT: Li, Yi
APPLICANT: Rosen, Craig A.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HPAJ70
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSER: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,
ADDRESSER: STUART & OLSTEIN
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: New Jersey
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,980
FILING DATE: 06-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Ferraro, Gregory D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-446
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1474 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 274..1233
US-08-465-980-1

alignment_scores:
Quality: 385.00 Length: 279
Ratio: 2.163 Gaps: 2
Percent Similarity: 63.799 Percent Identity: 31.900

alignment_block:
US-09-755-017-2 x US-08-465-980-1 ..

Align seq 1/1 to: US-08-465-980-1 from: 1 to: 1474
22 TrpLeuGlnPheProLeuLeuValAlaPheLeuIleSerThrThrValTh 38
   ::::::::::::::::::::
346 TGGGTTGGCTTCCCTCCCTTTCATG.....TATGTAGTGGC 383
38 rIlePheGlyAsnLeuThrIleIleLeuValSerArgLeuAspThrLysL 55
   ::::::::::::::::::::
384 AATCTGTGAACACTGCATCTGCTTTCATCGTAAGACGACGACGACGCC 433
55 euHisThrPrometThrPhePheLeuThrAsnLeuSerLeuLeuAspLeu 71
   ::::::::::::::::::::
434 TGCACGCTCCGATGTACTCTTCTCTGCAATGCTGACGCCATTTGACCTG 483
72 CysThrThrThrCysThrValProGlnMetLeuValAsnLeuCysSerI 88
   ::::::::::::::::::::
484 GCCTATCCACATTCACACATGCTAGATTCCTTGCCTTTTCTGCTTGA 533
88 eArgLysValIleSerThrArgLysCysValAlaGlnLeuPheIlePheL 105
   ::::::::::::::::::::
534 TTCCCGAGATATGACATGTAGGCGCTGTCTTACCCAGATGTCTTATATTC 583

```

```

105 euAlaLeuGlyAlaThrGluTyrLeuLeuLeuAlaValMetSerPheasp 121
    |||||:||||| |||:|||||:|||||:|||||
584 ATGCCCTTCAGCATTTGAATCACCATCTCTGCGCATGGCTTTGAC 633
    |||||:|||||:|||||:|||||:|||||:|||||
122 ArgPheValAlaIleCysArgProLeuHisTyrSerValIleMetHisgl 138
    |||||:|||||:|||||:|||||:|||||:|||||
634 CGTTATGTGGCATCTGCGACCCAGCTGCCCATGTGCTCAACAA 683
    |||||:|||||:|||||:|||||:|||||:|||||
138 nArgLeuCysLeuGlnLeuAlaIleAlaSerPValThrGlyPheSera 155
    :|||||:|||||:|||||:|||||:|||||:|||||
684 TACAGTAACAGCCAGATTGGCATGCTGCTGCTGCCGCGATCCCTCT 733
    |||||:|||||:|||||:|||||:|||||:|||||
155 snSerValTyrPleuSerThrLeuGlnLeuProLeuCysAspPro 171
    |||||:|||||:|||||:|||||:|||||:|||||
734 TTTTTCCTCCACTGCTGCTGATATACAGCGGTGCTTCTCCACTCC 783
    |||||:|||||:|||||:|||||:|||||:|||||
172 TyrValIleAspHisPheLeuCysGluValProAlaLeuLeuLysLeuS 188
    |||||:|||||:|||||:|||||:|||||:|||||
784 AATGCTCTCTGCGACTCTATTGTGTCCACACAGATGTATGAAGTTGGC 833
    |||||:|||||:|||||:|||||:|||||:|||||
188 rcysValGluThrThrAlaAsnGlnAlaGluLeuPheLeuValSerGluL 205
    :|||||:|||||:|||||:|||||:|||||:|||||
834 CTATGAGACACTTGGCCCATGTGTATATGCTTACTGCTCATCTCTGC 883
    |||||:|||||:|||||:|||||:|||||:|||||
205 eupHeHisLeuIleProLeuThrLeuIleLeuIleSerTyrAlaPheIle 221
    |||||:|||||:|||||:|||||:|||||:|||||
884 TGCTCATGGGCGGTGAGACGTATATCTCTCTGCTTCTTCTTGATAT 933
    |||||:|||||:|||||:|||||:|||||:|||||
222 ValArgAlaValLeuArgIleGlnSerAlaGluGlyArgGlnLysAlaPh 238
    :|||||:|||||:|||||:|||||:|||||:|||||
934 ATACGAAAGCTTCTGCAACTGCTCTCAAGTCAAGACGAGCGGCAAGGCTT 983
    |||||:|||||:|||||:|||||:|||||:|||||
238 eGlyThrCysGlySerHisLeuIleValIleSerLeuPheTyrSerThra 255
    |||||:|||||:|||||:|||||:|||||:|||||
984 TGGAACTGTGTGTACACATTGTGTGTGCTGCTTATGTATGTCAC 1033
    |||||:|||||:|||||:|||||:|||||:|||||
255 lavalSerValTyrLeuGlnProProSerProSerSerLysAspGlnGly 271
    :|||||:|||||:|||||:|||||:|||||:|||||
1034 TTTTGGCTCTCAGTTGTATACCGCTTTGGAAACAGCTTCTCATCCCAT 1083
    |||||:|||||:|||||:|||||:|||||:|||||
272 LysMetValSerLeu.....PheTyrGlyIleIleAlaProMetLeuAs 286
    |||||:|||||:|||||:|||||:|||||:|||||
1084 GTGCGTGTGTGATGAGTGATCTATCTGCTGCTGCTGCTGCTGATCA 1133
    |||||:|||||:|||||:|||||:|||||:|||||
286 nProLeuIleTyrThrLeuArgAsnLysGluValLys 298
    |||||:|||||:|||||:|||||:|||||:|||||
1134 TCCCATCATCTATGTGTGCCAAACAAACAGATCAGA 1170
    |||||:|||||:|||||:|||||:|||||:|||||
seq_name: /cgn2_6/ptodata/2/jna/5B_COMB.seq:US-09-053-303-1
seq_documentation_block:
; Sequence 1, Application US/09053303
; Patent No. 5948890
; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel R.
; APPLICANT: Li, Yi
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HPRAJ70
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAUM, GILFILLAN, CECCHI,
; ADDRESS: STUART & OLSTEIN
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/053.303
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/465,980
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Ferraro, Gregory D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-446
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1474 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 274..1233
US-09-053-303-1

alignment_scores:
Quality: 385.00 Length: 279
Ratio: 2.163 Gaps: 2
Percent Similarity: 63.799 Percent Identity: 31.900

alignment_block:
US-09-755-017-2 x US-09-053-303-1 ..
Align seq 1/1 to: US-09-053-303-1 from: 1 to: 1474

22 TyrLeuGlnPheProLeuLeuValAlaPheLeuIleSerTyrThrValTh 38
    |||||:|||||:|||||:|||||:|||||:|||||
346 TGGGTGGCTTCCCTCCCTTCATG.....TATGTATGGC 383
    |||||:|||||:|||||:|||||:|||||:|||||
38 rIlePheGlyAsnLeuThrIleLeuValSerArgLeuAspThrLysL 55
    :|||||:|||||:|||||:|||||:|||||:|||||
384 AATGTGTGGAACATGCTGCTGCTTCATCTGAGACGACGACGACGCC 433
    |||||:|||||:|||||:|||||:|||||:|||||
55 euHisThrProMetTyrPhePheLeuThrAsnLeuSerLeuLeuAspLeu 71
    |||||:|||||:|||||:|||||:|||||:|||||
434 TGCACGCTCCGATGTACCTCTTCTCTGATGCTTGCAGCGCATGTGAC 483
    |||||:|||||:|||||:|||||:|||||:|||||
72 CysTyrThrThrCysThrValProGlnMetLeuValAsnLeuCysSerI 88
    :|||||:|||||:|||||:|||||:|||||:|||||
484 GCCTTATCCACATCCACCATGCTTAAGATCTTGCCTTTCTGGTTTGA 533
    |||||:|||||:|||||:|||||:|||||:|||||
88 eArgLysValIleSerTyrArgGlyCysValAlaIleLeuPheIlePhe 105
    :|||||:|||||:|||||:|||||:|||||:|||||
534 TTCCCGAGAGATTAAGCATTTGAGGCGTCTTACCAGATTTCTTTATTC 583
    |||||:|||||:|||||:|||||:|||||:|||||
105 euAlaLeuGlyAlaThrGluTyrLeuLeuLeuAlaValMetSerPheasp 121
    |||||:|||||:|||||:|||||:|||||:|||||
584 ATGCCCTTCAGCATTTGAATCACCATCTCTGCGCATGGCTTTGAC 633
    |||||:|||||:|||||:|||||:|||||:|||||
122 ArgPheValAlaIleCysArgProLeuHisTyrSerValIleMetHisgl 138
    |||||:|||||:|||||:|||||:|||||:|||||
634 CGTTATGTGGCATCTGCGACCCAGCTGCCCATGTGCTCAACAA 683
    |||||:|||||:|||||:|||||:|||||:|||||
138 nArgLeuCysLeuGlnLeuAlaIleAlaSerPValThrGlyPheSera 155
    :|||||:|||||:|||||:|||||:|||||:|||||
684 TACAGTAACAGCCAGATTGGCATGCTGCTGCTGCCGCGATCCCTCT 733
    |||||:|||||:|||||:|||||:|||||:|||||
155 snSerValTyrPleuSerThrLeuGlnLeuProLeuCysAspPro 171
    |||||:|||||:|||||:|||||:|||||:|||||
734 TTTTTCCTCCACTGCTGCTGATATACAGCGGTGCTTCTCCACTCC 783

```

```

172 TyrValIleAspHisPheLeuCysGluValProAlaLeuLeuLysLeu 188
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
784 AATGTCCTCGCACCTCTATGTGTGTCCACAGATGATATGAAGTTGCC 833
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
188 rcysValGluThrAlaAsnGluAlaGluLeuPheLeuValSerGlu 205
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
834 CTATGCAGACCTTTGCCCAATGTGTATGTGCTTACTACCACTTCTGC 883
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205 euphHisLeuIleProLeuThrLeuIleLeuIleSerTyrAlaPheIle 221
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
884 TGCTCATGGCGGTGACGATATGTATCTCTTGTCTTATTTCTGATA 933
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 ValArgAlaValLeuArgIleGlnSerAlaGluGlyArgGlnLysAla 238
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
934 ATACGAAGGCTTGCACACTGCGCTTCCAACTCAGAGCGGCGCAAGCGCT 983
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
238 eGlyThrCysGlySerHisLeuIleValValSerLeuPheTyrSerThra 255
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
984 TGGAACTGTGTGTACACATGTGTGTGTCTGCCCTTCTATGTGCCAC 1033
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
255 ValAlaSerValTyrLeuGlnProProSerProSerSerLysAspGlnGly 271
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1034 TTATGTGCGCTCTCAGTTGTACACCGCTTTGGAACACCTTCATCCCAT 1083
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
272 LysMetValSerLeu.....PheTyrGlyIleIleAlaProMetLeuAs 286
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1084 GTCCGTGTGTGTACATGGGTGACATCTACCTGTCTGCGCTCTCATCA 1133
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
286 nProLeuIleTyrThrLeuArgAsnLysGluValLys 298
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1134 TCCCATCATCTATGTGTGCCCAACCAACAGATCAGA 1170
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: /cgn2_6/ftodata/2/1na/PCtUS_COMB.seq: PCT-US95-07093-1
seq_documentation_block:
; Sequence 1, Application: PC/TUS9507093
; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel R.
; APPLICANT: Li, Yi
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HPR470
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,
; ADDRESSER: STUART & OLSTEIN
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07093
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/465,980
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferraro, Gregory D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-446
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 1474 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 274..1233
; PCT-US95-07093-1

alignment_scores:
    Quality: 385.00    Length: 279
    Ratio: 2.163      Gaps: 2
    Percent Similarity: 63.799    Percent Identity: 31.900

alignment_block:
US-09-755-017-2 x PCT-US95-07093-1
Align seg 1/1 to: PCT-US95-07093-1 from: 1 to: 1474

22 TrpLeuGluPheProLeuLeuValAlaPheLeuIleSerTyrThrValTh 38
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
346 TGGGTGGCTCCCTCCCTTCCATG.....TATGTAGTGGC 383
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
38 rIlePheGlyAsnLeuThrIleIleLeuValSerArgLeuAspThrLysL 55
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
384 AATGTGTGGAACTGCATGCTGTGCTTTCATCTGTAAGACGACGACGACG 433
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
55 euhIsthrProMetTyrPhePheLeuThrAsnLeuSerLeuLeuAspLeu 71
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
434 TGCACCTCCGATGATCTCTCTCTGATGCTTGCACGACCATGTGACCTG 483
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
72 CysTyrThrThrCysThrValProGlnMetLeuValAsnLeuGlySerIle 88
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
484 GCCTTATCCACATCCACCATGCTTAAGATCTTGCCCTTTGTGTTTGA 533
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88 eArgLysValIleSerTyrArgGlyCysValAlaGlnLeuPheIlePheL 105
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
534 TTCCCGAGAGATTTAGCATTTGAGCGCTGTCTTACCAGATTTCTTTATTC 583
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
105 euhAlaLeuGlyAlaThrGluTyrLeuLeuLeuAlaValMetSerPheasp 121
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
584 ATCCCTCTCAGCATTTAGTCCACCATCATCTCTGCTGCGCATGCTTGTGAC 633
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122 ArgPheValAlaIleCysArgProLeuHisTyrSerValIleMetHisG 138
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
634 CGTTATGTGGCATGTGCCACCCACGCGGCATGTGCACATGCTCAACAA 683
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
138 nArgLeuCysLeuGlnLeuAlaAlaIleSerTrpValThrGlyPheSerA 155
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
684 TACAGTAACAGCCACAGATTGGCATCTGCTGTGTGTCGGGATCCCTCT 733
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
155 snSerValTrpLeuSerThrLeuThrLeuGlnLeuProLeuCysAspPro 171
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
734 TTTTTCCTCCACGTCTGCTGATGATCAAGGCGCTGCTGCACTCC 783
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
172 TyrValIleAspHisPheLeuCysGluValProAlaLeuLeuLysLeu 188
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
784 AATGTCCTCGCACCTCTATGTGTGCCACAGAGATATGATGAGTTGGC 833
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
188 rcysValGluThrThrAlaAsnGluAlaGluLeuPheLeuValSerGlu 205
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
834 CTATGCAGACCTTTGCCCAATGTGTATGTGCTTACTACCACTTCTGC 883
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205 euphHisLeuIleProLeuThrLeuIleLeuIleSerTyrAlaPheIle 221
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
884 TGCTCATGGCGGTGACGATATGTATCTCTTGTCTTATTTCTGATA 933
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 ValArgAlaValLeuArgIleGlnSerAlaGluGlyArgGlnLysAla 238
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
934 ATACGAAGGCTTGCACACTGCGCTTCCAACTCAGAGCGGCGCAAGCGCT 983
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

238 eglYthrcysglyserHisleuIleValIserLeuPheYrSerThra 255
|||||  |||||  |||||  |||||
984 TGAACCTGTGTGCACACATTGTGTGTCTATCCCTTCTATGTGCCAC 1033
255 IalSerValTyrLeuGlnProPserProSerSerLysaspGlnGly 271
|||||  |||||  |||||  |||||
1034 TTATGGCCTCTCAGTTGTACACCGCTTGGAAACAGCCTTCATCCATT 1083
272 LysMetValSerLeu.....PheYrGlyIleIleIalPrometLeuAs 286
|||||  |||||  |||||  |||||
1084 GTGCGTGTGTGCATGGTGTACATCTACCTGTGCTGCTCTGTATCAA 1133
286 nProleuIleTyrThrLeuArgAsnLysGluValLys 298
|||||  |||||  |||||  |||||
1134 TCCATCATCTATGTGTGCCAAACCAACCAACAGATCAGA 1170

```

```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-599-252-84

```

```

seq_documentation_block:
; Sequence 84, Application US/08599252
; Patent No. 5705343
;
; GENERAL INFORMATION:
; APPLICANT: DRAVNA, DENNIS T.
; APPLICANT: FEDER, JOHN N.
; APPLICANT: GNIKE, ANDREAS
; APPLICANT: KIMMEL, BRUCE E.
; APPLICANT: THOMAS, WINSTON J.
; APPLICANT: WOLFF, ROGER K.
; TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY
; TITLE OF INVENTION: HEMOCHROMATOSIS
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Suite 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/599,252
; FILING DATE: 09-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 9053-0001.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-0763
; TELEFAX: (202) 887-1500
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-599-252-84

```

```

alignment_scores:
Quality: 227.00      Length: 68
Ratio: 3.721        Gaps: 0
Percent Similarity: 89.706      Percent Identity: 60.294

```

```

alignment_block:
US-09-755-017-2 x US-08-599-252-84/rev ..
Align seg 1/1 to reverse of: US-08-599-252-84 from: 1 to: 1320

```

```

240 ThrCysGlySerHisLeuIleValIserLeuPheYrSerThraIaVa 256
|||||  ||||||  ||||||  ||||||  ||||||  ||||||  |||
1319 ACTGCAGCTGCATCTATATGTGTCTCTCTCTCTCTCTCTCTCTCTCT 1270
256 IserValTyrLeuGlnProPserProSerSerLysaspGlnGlyLys 273
|||||  ||||||  ||||||  ||||||  ||||||  ||||||  |||
1269 ATACATGTACCTCCAGCCAGCAATACTTATCCAGGAGCCAGGGCAGT 1220
273 etValSerLeuPheYrGlyIleIleIalPrometLeuAsnProleuIle 289
|||||  ||||||  ||||||  ||||||  ||||||  ||||||  |||
1219 TCTTACCTCTTCTACACAATGTCTACCTCCAGCTTAACCCCTGATC 1170
290 TyrThrLeuArgAsnLysGluValLysGluGlyPheLysArgLeuValAl 306
|||||  ||||||  ||||||  ||||||  ||||||  ||||||  |||
1169 TATACACTAGAAACAAAGATGTAAAGAGGCCATGAGAAAGTGCTAGG 1120
306 ARG 307
|||
1119 GAAG 1116

```

```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-436-074-57

```

```

seq_documentation_block:
; Sequence 57, Application US/08436074
; Patent No. 5753438
;
; GENERAL INFORMATION:
; APPLICANT: DRAVNA, DENNIS T.
; APPLICANT: FEDER, JOHN N.
; APPLICANT: GNIKE, ANDREAS
; APPLICANT: KIMMEL, BRUCE E.
; APPLICANT: THOMAS, WINSTON J.
; APPLICANT: WOLFF, ROGER K.
; TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY
; TITLE OF INVENTION: HEMOCHROMATOSIS
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Suite 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,074
; FILING DATE: 08-MAY-1995
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 9053-0001.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-0763
; TELEFAX: (202) 887-1500
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-436-074-57

```

```

alignment_scores:
Quality: 227.00      Length: 68
Ratio: 3.721        Gaps: 0
Percent Similarity: 89.706      Percent Identity: 60.294

```

alignment\_block:  
US-09-755-017-2 x US-08-436-074-57/rev ..

Align seg 1/1 to reverse of: US-08-436-074-57 from: 1 to: 1320

```
240 ThCysGlySerHisLeuIleValSerLeuPheTyrSerThrAlaVal 256
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1319 ACTTGAGCTCGCATCTAATTGTTGCTCTCTCTATGTCATCAATCAT 1270
256 lSerValTyrLeuGlnProProSerProSerSerLysAspGlnGlyLysM 273
: |||||:|||||: |||||: |||||: |||||: |||||: |||||:
1269 ATACATCTACTCCAGCCAGCAATACTTATTCCTCCAGACCAAGGCACAT 1220
273 eValSerLeuPheTyrGlyIleIleAlaProMetLeuAsnProLeuIle 289
: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
1219 TTCTTACCCCTTTTCTACACAATGTGTACCTCCAGTGTAAACCCCTGATC 1170
290 TyrThrLeuArgAsnLysGluValLysGluGlyPheLysArgLeuValAl 306
|||||:|||||: |||||: |||||: |||||: |||||: |||||:
1169 TATACACTAAGAAACAAGATGTTAAAGAGGCATGAAAGAGGTGCTAGG 1120
306 aArg 307
: |||||:
1119 GAAG 1116
```